

British science sent back to dog-house

Not for the first time, a British government has revealed its lack of understanding of its scientific enterprise by means of an insensitive reorganization forced by reasons having nothing to do with research.

BRITISH governments do not regard consistency as a virtue, but feel an almost moral obligation to behave differently from their predecessors, and are not ashamed to throw their previous decisions to the winds if convenience so dictates. That is what happened last week with the British government's new arrangements for the administration of the civil science budget (see page 103). At the stroke of a pen or two, British science has gone from being at the "heart of government" (to coin a phrase), with its administration in the hands of the Office of Science and Technology (OST) and its own representative in the Cabinet in the quaintly named Chancellor of the Duchy of Lancaster, to being just another branch of the Department of Trade and Industry (DTI).

Pure malice is probably not the reason why the appletart has been overturned, but it cannot be excluded. It would be more in the tradition of the failure of successive British governments to appreciate the role and sensitivity of science in their affairs that last week's upheaval should have been driven by considerations of administrative tidiness, false as they turn out to be. But whatever the motives, the changes are potentially disastrous.

No doubt in haste to complete a government reorganization made necessary by the Prime Minister's re-election as head of the Conservative Party, those concerned have overlooked the intentions they made public in the 1993 white paper *Investing in our future*. Then, the same government said that OST would play a part in coordinating and stimulating research throughout government, the DTI included. It was understood that success in this regard would depend on OST's competence, the intellectual force of its arguments and, crucially, the backing of a single-minded minister and even of the Prime Minister.

How does anybody think that possible now that OST has become merely part of one of the government departments whose better articulation with other departments is supposed to be part of its role? The charitable view is that the Prime Minister and Mr Michael Hesletine, now his deputy, have simply overlooked this part of their plan. But Professor John Cadogan, director-general of the research councils, should know better. His promise that it will be "business as usual" for OST can be taken seriously only he implies that its coordination role will be quietly abandoned.

Several worries now arise. How will fundamental sciences such as particle physics and radioastronomy, which cannot even pretend to pay their way by contributing to

wealth creation, justify continued support in the new setting? Can it be in the public (or even the government's) interest that scrutiny of decisions on science in the House of Commons will be as attenuated as now seems likely (see page 103)? To whom will the new Science Adviser to the government, Professor Robert May, have access in the new regime and over whom, and what, will he have authority? It is to be hoped that answers to questions like these will be wrung from the government before its thoughtless reorganization is set in stone.

But no amount of explanation will remedy the chief damage that has been done. The British research community has accepted with a good grace the arrangements introduced three years ago because they promised openness, a sense of partnership between the community and government and the promise that the whole of government science would eventually be judged by the tests proposed for the research councils. A delicate balance, it nevertheless has brought about an improvement of morale in research laboratories everywhere. That is now in hazard. More than mere fast talking will be required to make sure that good spirits do not melt away as quickly as they returned.

The government, of course, will say that these fears are exaggerated. Is not science better off with Mr Michael Hesletine, the most powerful minister in the government, as the ultimate overlord? That is what was said when Mrs Margaret Thatcher (as she was), trained in chemistry at Oxford, was prime minister. She promised to pay personal attention to research, but in the event found she had not the time. Will Hesletine, with all the government's cares apparently on his shoulders, have more time for science? That seems as likely as that it will "business as usual" at OST. □

Physics not guilty

The 50th anniversary next week of the first nuclear test impells anticipation of the outcome of innovation.

FIFTY years ago next Monday, 17 July, the group of people gathered before dawn at a remote site in the New Mexico desert knew they were to witness an event that would change the world: the first explosion of a nuclear bomb. Scattered in bunkers 10,000 yards from the designated explosion point (called "Ground Zero"), the watchers included many who shaped the development of science in



the West in the succeeding generation. Bainbridge, Bethe, Conant, Fermi, Kistiakowsky, von Neumann, Penney and Peierls (two Brits), Rabi, Ramsey, Serber, Teller, Weisskopf — this partial list of those who had helped make the first nuclear device explode is a good substitute for a roster of those who had or, afterwards, would make post-war physics ascendant for at least the two decades ahead. Their leader, J. Robert Oppenheimer, the director at Los Alamos since its foundation in 1943, was characteristically ambivalent when the explosion took place as planned. He shared his colleagues' delight at their achievement, but he agonized about the consequences. "The physicists have known guilt", he said, in an oblique reference to the legend of Prometheus and to the fear that people can know too much for their own good.

The parallel is too evocative to be forgotten, but is it real? Within its terms of reference, the Manhattan Project was a huge success. The Los Alamos part of it, where the first and succeeding bombs were designed and built, was a staggering feat. The laboratory had been founded only in the spring of 1943 with the objective of designing a workable bomb and of carrying out the research required to prove and test the design. Fissile material was to be made elsewhere (Oak Ridge for uranium and Hanford for plutonium). Oppenheimer's solution was to recruit the brightest and the best from wherever he could. Feynmann was also one of them, Niels Bohr was for much of the time the father-figure in residence at Los Alamos (but had returned to Copenhagen by the time of Trinity). It is inconceivable that Oppenheimer could have persuaded this galaxy of *prima donnas* to their hum-drum tasks without his distinctive intelligence and even his ambivalence. The question of how to prevent the arms race that would follow the first explosion, repeatedly put to him by Szilard and, more influentially, by Bohr, was constantly discussed. That ambition came to nothing, as we know. But Trinity remained for many years a cheerful sign that anything is possible if enough intelligence and skill are devoted to it.

Was it also the promethean cataclysm Oppenheimer foresaw? The import of the legend conflicts with the contrary opinion that knowledge, as such, is unalloyed enlightenment, and that what matters is the use people make of what they know. It is relevant that well-placed Oppenheimer had earlier argued against the bombing of Japan, urging a demonstration of the power of the bomb instead. That option was quickly closed by the diplomats and the military, who had put Hiroshima at the head of a list of targets in Japan by the time of Trinity. Earlier in 1945, Bohr had put to President Franklin D. Roosevelt his belief that Trinity and what would follow it would be so obviously terrible that it could be used to terrify governments into international collaboration to avoid an arms race and had struck a chord with Roosevelt, who then sadly and inconsiderately died. Oppenheimer went on to argue to the US government that it should declare its willingness to share its knowledge of the peaceful benefits of nuclear fission with all-comers, much in the terms of President Dwight D. Eisenhower's "Atoms for Peace" programme a decade later, but the diplomats, still

smarting from the partition of Europe they had agreed at the Yalta summit the previous year, would have none of it; Oppenheimer's advocacy became a stick with which he was beaten a few years later at his appeal against the decision of the US Atomic Energy Commission to deny him his security clearance.

With the benefit of hindsight, it is plain that Los Alamos was far from single-throated in its opinions of what should follow Trinity. Some shared the view, common then in the United States, that everything possible should be done to shorten the Pacific War. With US casualties in the capture of the barren volcanic strip called Iwo Jima fresh in people's minds, who could call that unnatural? Given that urgency, together with the possibility that the first explosion might be a fizzle or that its yield might be much less than expected, there may well have been no other course than to drop the first bomb in anger. In the event, the Hiroshima bomb did not bring about the "unconditional surrender" demanded by the Potsdam summit a few days earlier, the meeting at which Truman (the new US president) told Stalin of a weapon of unprecedented destructive power. (Truman did not know that Stalin already knew what the Manhattan project was all about.) The guess is probably correct that if Truman had gone on to put Oppenheimer's argument to Stalin, he would have found that it cut little ice. If only the Pacific War had ended before Trinity became possible, the course of recent history may have been different. But playing history games is a fruitless pursuit.

Two things nevertheless stand out from the period of the great arms race, 1945-90, now ended. First, physicists (many from Los Alamos itself) played important parts in helping to alert public opinion to the dangers of the arms race. For what it is worth, Oppenheimer's plea for openness was vindicated, a few months after Hiroshima, by the publication of the Smyth report describing in detail and plain language what fission was about. By 1946, the United States had also urged on the United Nations (and was rebuffed, chiefly by the Soviet Union and its allies) a plan for the international regulation of the uses of nuclear energy, giving some attention to Oppenheimer's plan.

Second, it is plain that dealings between the research community, represented by those gathered at the Trinity site 50 years ago, and the government were those between servants (however distinguished) and masters. The servants did the work and the masters literally called the shots or, as British ministers used to say (and may have done again — see below), "science should be on tap, not on top". The truth is that the Los Alamos crew included people with a more accurate vision of where Trinity might lead than their political masters, but were less able to argue their case effectively. Their guilt was not promethean guilt, which cannot be. If they had guilt at all, it lay in their readiness to accept the servile role with which their masters were most comfortable. But what can that matter, with the Cold War now ended? Because, for one thing, a quarter of a million people were killed at Hiroshima and Nagasaki. And, for another, the dilemma will recur. Will the research community be ready when it does? □

Concern and anger greet shift of UK science unit into industry agency

London. Britain's scientific community experienced a mixture of bemusement and anger last week after a series of top-level government changes announced by John Major, the prime minister, in the immediate aftermath of his successful re-election as leader of the Conservative Party.

At the centre of the changes has been Michael Heseltine, previously president of the Board of Trade, and as such, head of the Department of Trade and Industry (DTI). He has been promoted to deputy prime minister in recognition of his support for Major's successful re-election.

Heseltine, who has long argued for science to be integrated into industrial and economic policy, will have broad-ranging powers in his new post. These will, in particular, include the chairmanship of the cabinet's science and technology subcommittee, made up of all cabinet ministers with an interest in research.

He will also take with him to the Cabinet Office members of the DTI's team concerned with the international competitiveness of British industry, responsible for a white paper on the topic published two months ago in conjunction with the results of the government's technology foresight exercise (see *Nature* 375, 265; 1995).

But at the same time, the Office of Science and Technology (OST), which is responsible for the work of Britain's seven research councils, will move in the opposite direction. Set up three years ago by Major within the Cabinet office as part of the Office of Public Service and Science, the OST is now being shifted to the DTI.

As a result, Britain's £1.2-billion (US\$1.9-billion) annual science budget will now become the direct responsibility of Heseltine's successor, Ian Lang. So, too, will the OST's formal role in co-ordinating the research and development activities of all government departments.

The move provoked an immediate outcry from parts of the scientific community, which had been hoping for a period of stability after the upheavals that followed the white paper of May 1993. Many scientists claim that allocating responsibility for science policy to a government department concerned primarily with promoting industrial innovation is likely to increase the difficulties of securing support for long-term fundamental research.

There is also concern that the move will inevitably undermine efforts over the past three years to use the OST's central position within the Cabinet Office to increase the impact and awareness of science in all spheres of government activity.

"The big danger is that we are back to the situation where science and technology, which should be part of across-the-board policy in government, are reduced to a junior role in another ministry with very much larger responsibilities," says Denis

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Spot the minister: Ian Lang (right) will be responsible for science, but Michael Heseltine (left) will wield influence from the centre.

Noble, professor of cardiovascular physiology at the University of Oxford, and co-founder of the group Save British Science.

The concerns of scientists were only increased by a press release from Downing Street. This justified the move on the grounds that it "will allow the government's policy on science, engineering and technology to be developed alongside its policies on industry, and with due regard to the contribution of science, engineering and technology to long-term wealth creation."

Many were quick to point out, for example, that this statement makes no reference to a complementary goal which has — at least up to now — accompanied statements of the government's strategy for science, namely that it should also contribute to the

quality of life. And that, even with the qualification 'long-term', it seems to suggest that research policy will in future be dominated by the needs of British industry.

"This is one step that the government will come to regret bitterly" predicts David Triesman, general secretary of the Association of University Teachers, the body which represents the interests of academic staff. "Reducing science merely to a hand servant of industry is madness."

Government officials have been playing down the implication of the shift by emphasizing that the OST will still operate as a single unit within the DTI. They also emphasize that Robert May, who is about to take over as the head of the office, will still be chief scientific adviser to the government, with direct access to the prime minister.

In an attempt to allay concerns, Sir John Cadogan, director-general of research councils, has written to the heads of all the research councils, pointing out that the functions of the OST — including the budgets of the research councils — will be "ring-fenced" within the DTI, and that he will still report directly to the cabinet minister formally responsible for science, now Lang.

In his letter, Cadogan also points out that the OST will continue to exercise responsibility for activities such as the Technology Foresight programme and the LINK programmes. "The message therefore is business as usual for the OST", wrote Cadogan.

Some scientific bodies have been cautious in their response. An official at the Royal Society, for example, said that the society was taking time to consider how it was going to react, and that it wanted to see the ►

Parliamentary scrutiny set to fall

London. One of the side-effects of last week's decision by the British government to merge the Office of Science and Technology (OST) into the Department of Trade and Industry (see above) is likely to be a reduction in the Parliamentary scrutiny of science policy decisions.

In particular, the move could mean the disappearance of the House of Commons Select Committee on Science and Technology, the committee which currently holds broad oversight responsibility for government decisions related to science.

It may also lead to the disappearance of the protected spot which science issues have, for the past few years, enjoyed during question time in the House of Commons, which allows Members of Parliament (MPs) to quiz the minister for

science in person on topics of their choice.

Since 1979, the responsibilities of select committees have been aligned with individual government departments. This initially meant that science — which had previously enjoyed its own committee — was lumped with education and the arts under the committee overseeing the Department of Education and Science.

The creation of the OST in the wake of the 1992 election meant that, for the first time in thirteen years, science and technology had their own dedicated select committee. Under the chairmanship of Sir Giles Shaw, Conservative MP for Pudsey, this has actively pursued a number of wide-ranging inquiries.

The most recent has been a 12-month investigation, including numerous ►

► details of the new arrangements rather than passing any hasty judgement.

But others have greeted Cadogan's statement with scepticism, pointing out that there is little in the DTI's recent track record to suggest it is genuinely committed to long-term fundamental research. Many have been quick to point out, for example, that since the publication of the white paper on science in May 1993, the research and development budget of the department has fallen substantially.

Admittedly, much of the reduction has taken place in technology development programmes, rather than basic research, and the DTI has built strong links with various research councils, in particular the Engineering and Physical Sciences Research Council. But the department, under Heseltine's leadership, also abolished the post of chief scientific adviser shortly after the publication of the white paper.

"We note that the OST is going to be ring-fenced within the DTI, but it will be a ring fence in the middle of a lion compound," says Alun Jones, chief executive of the Institute of Physics. "It makes sense for science to be linked closer to technology and industry. But we are disappointed that it appears to have moved lower down the government's priorities."

Unsurprisingly, the decision to move the OST into the DTI has come under fierce criticism from the opposition Labour Party, which itself considered — but rejected — proposing such a move as part of its own science policy two years ago. John Battle, Labour's science spokesman, says that the move represents "the demotion of science and technology".

Meanwhile in the scientific community itself, perhaps the greatest immediate concern is over the future of government-funded laboratories owned and run by the research councils.

Although moves to 'privatize' organizations such as the Medical Research Council's Laboratory of Molecular Biology in Cambridge had previously been headed off, the appointment of Heseltine, a keen advocate of this strategy, suggests that such proposals are likely to come in for renewed attention.

David Dickson

► hearings as well as a visit to the United States, of the implications of human genetics. Its report, due next week, is expected to contain a wide sweep of recommendations, from amendments to patent legislation limiting the scope of patents on human genes, to the need for regulations on the use of genetic information by insurance companies.

If Parliament follows its established rules of procedures — and rejects moves currently under consideration to depart from its accepted practice — all such issues will now be dealt with by a committee whose main brief will be to look at all the activities of the DTI, ranging

US may drop federal support for families of researchers

San Francisco. The Clinton administration has proposed eliminating rules that allow research universities to put the relatives of faculty members through college at the taxpayer's expense. Universities say that they need the benefit in order to attract and retain top researchers. But critics say the government gets nothing in return.

According to the General Accounting Office (GAO) — the investigatory arm of Congress — between 1991 to 1993, four top US research universities charged \$17 million to the government for the tuition costs of faculty family members. The GAO studied five institutions: the Massachusetts Institute of Technology, Stanford University, the University of California, John Hopkins University and the University of Chicago.

MIT charged the highest portion — 56 per cent — of its programme costs of \$12,566 per family member to federal research contracts and grants. The University of Chicago, which has the most generous programme, charged the least, with 15 per cent of its \$16,842 in costs repaid by the government. The University of California asked for no reimbursement; it does not provide financial assistance to relatives of faculty, on the grounds that its tuition charges are comparatively low.

The GAO found that education benefits had been a long-standing tradition at the universities since the 1960s. The practice was initially intended to help overcome a shortage of scientists, engineers and other researchers, says Charles Thompson, assistant director in acquisition policy, technology and competitiveness issues for the GAO. "Today, the environment may be different," he adds, pointing out that the GAO did not recommend that the programme should be suspended — merely that it should be examined more closely.

Universities say the benefit remains an important recruiting tool as the salaries they offer continue to lag far behind those in the

commercial sector. They argue that the government should allow universities to manage their overall compensation programmes, as long as the total amount of money is reasonable. Cutting the government reimbursement policy would only force them to offer higher salaries which would then be charged back to grants, universities add.

But the Clinton administration, after one look at the GAO report, proposed the reimbursement be slashed. "It was our judgment there wasn't any real benefit coming back to the government itself," says Lawrence Haas, associate director for communications at the Office of Management and Budget. He said universities may need the benefit, but that doesn't mean the government should have to pay for it.

Haas said his office would carefully consider the 200 or so responses it had received and was planning to issue a final decision within a couple of months. **Sally Lehrman**

France urged to set up genetics network

Paris. France's healthcare system has failed to adapt adequately to advances in genetics research, and urgently needs to create a national network of specialized genetic clinics staffed by physicians trained in medical genetics.

These are the main conclusions of a report submitted last week to Elisabeth Hubert, the minister of health, by Jean-François Mattei, who is both head of the Paediatrics and Medical Genetics Department at the Timone hospital in Marseille, and a member of the National Assembly (UDF, Bouches-du-Rhône).

The report claims that consultations in clinical genetics are haphazard, and are often only available from research groups at major teaching hospitals. Indeed, Mattei points out that medical genetics was not recognised as a speciality until a law making it one was passed earlier this year.

To address these deficiencies, Mattei recommends that one clinical genetics centre be set up for every 300,000 inhabitants. Each centre would include a clinical geneticist, a cytologist, and a molecular biologist.

Mattei argues that such a network will also help to regulate the development of medical genetics, and in particular avoid the setting up of private centres offering genetic testing. Support for his recommendations is likely to come from the government. Alain Juppé, the new prime minister, called for a national programme in medical genetics in a speech made shortly after his nomination in May.

Declan Butler

from the management of pension funds to the operation of frontier customs. Furthermore, any questions to ministers on science-related topics could now end up being handled by a junior minister within the DTI.

Both moves have already raised concern among professional bodies such as the Royal Society of Chemistry which have grown to appreciate the opportunity that the existence of the science select committee — as well as that of a protected 'science questions' time — has provided them and others to air opinions and grievances in a formal parliamentary setting.

D. D.

German students to pay for funding boost...

Munich. The German government last week proposed a budget which would effectively give the ministry for education and research (BMBF) a 2.3 per cent funding increase next year, with basic research enjoying a 3 per cent growth. These figures contrast with a suggested reduction of 1.3 per cent in overall public spending, the first country's first budget reduction since 1953.

But only part — 0.6 per cent, from DM15.53 billion (US\$11 billion) to DM15.62 billion — of this proposed growth would come as cash. The rest would be met out of savings in the cost of government-sponsored student loans, proposed in a controversial bill drawn up by research minister Jürgen Rüttgers, which would introduce for the first time interest charges on loans.

The opposition social democrat party (SPD), which has pledged to increase support for research and technology by 10 per cent if elected to power, says that it will reject the bill when it goes before the Bundesrat, Germany's second chamber made up of representatives of the 16 *Länder*, where the SPD has a majority.

Peter Glotz, the party's education and research spokesman, says that Rüttgers is asking students to pay for the building-up of his ministry. He has called instead for a broad debate on the future financing of education, addressing the needs for a basic reform of a system which can no longer support a recent spurt in student numbers.

Changes to the student loan system — known as BAföG, an abbreviation of *Bundesausbildungsförderungsgesetz* — are a sensitive topic. Universities in particular have reacted angrily to the BAföG bill, described by Hans-Uwe Erichsen, president of the University Rectors Conference, as "totally unacceptable". This view has wide public support. The influential newspaper, the *Süddeutsche Zeitung*, described Rüttgers' juggling of the accounts as "a cheap trick".

But Rüttgers has tried to deflect such criticism by promising that the DM1.6 billion that could be raised over the next five years by requiring interest to be paid on loans would be used to increase support for universities and science — a move which he hopes his critics will approve of.

In particular, Rüttgers wants some of this money to be used for university building and equipment. Although the *Länder* are responsible for the running costs of their own universities, capital investment is shared equally between federal and *Länder* governments.

This system came under strain this year when the federal government put a cap on its share of investments which the *Länder* say is unreasonably low (see *Nature* 353, 95; 1995). Rüttgers clearly hopes the prospect of this cap being raised may make the *Länder*

drop their resistance to his BAföG bill. University building and special university research programmes are together marked for a 2.5 per cent increase in 1996.

According to Rüttgers' plan, both basic and applied research would benefit from his proposed funding changes, being earmarked for a 3 per cent and 3.8 per cent rise respectively. Two particular beneficiaries would be the Max Planck Society, which runs basic research institutes, and the Deutsche Forschungsgemeinschaft, to both of which Rüttgers has promised a 5 per cent increase if the BAföG bill is passed (see below).

More significantly, he has also promised to maintain this rate of increase for each of the following four years, to provide a constantly growing budget. But whether this will materialize depends on the agreement of the *Länder* governments, which co-finance the two organizations.

Rüttgers is also proposing to increase government spending on biotechnology, information technology and environmental research, in line with promises made in his first statement on research policy since he took office last November (see *Nature* 373, 648; 1995).

Funding of space research will not increase under his proposed budget — but neither will it decrease. Rüttgers says that "the BMBF thereby underlines its role as a reliable partner in, and ensures finances for, a realistic concept of European space programmes".

The BAföG bill will be discussed in the Bundesrat after summer. If it is rejected, as many are currently predicting, Rüttgers will be left with a budget rise that will be substantially below inflation. He refuses to comment on what his priorities would be in such a case.

Alison Abbott

...as scientists hope for the best

Bonn. "Politicians have discovered the 'future factor'", says Wolfgang Frühwald, president of the Deutsche Forschungsgemeinschaft (DFG), the organization which funds most university research in Germany. "That is why we are predicting a rise in research funds in next year's budget".

Frühwald is particularly optimistic about the renewal of the so-called five-by-five agreement, under which both the DFG and the Max Planck Society had been guaranteed a five per cent increase in budget for the five years 1991 to 1995.

Last week's promise by the government to renew its side of the agreement until 1999, provided parliament approves a controversial bill to impose interest on student loans (see above), supports his optimism.

Speaking at the annual meeting of the DFG two weeks ago, Frühwald said that a guaranteed increase of five per cent a year is needed as a basis for future policy, particularly given the extra responsibilities which the DGF shares with the Max Planck Society for research in the new *Länder*.

But the renewal of the five-by-five agreement does not depend on Bonn, as financing for both organizations is

shared equally between federal and *Länder* governments, and finance ministries in some *Länder* have not been enthusiastic about such a major commitment.

The *Länder* are expected to make their final decision after the summer. But several influential figures spoke in favour at the DFG annual meeting. They included Johannes Rau, prime minister of Nordrhein-Westfalen, and Hans Zehetmair, science minister of Bavaria.

According to Frühwald, another optimistic sign is the support from Anke Brunn, deputy chairwoman of the Bund-Land Commission (BLK), the body which brokers decisions on education and science between federal and *Länder* governments.

A renewal of the agreement would mean an additional DM447 million for the DFG, which is already planning how best it could be spent. Two weeks ago, the DFG opened up talks with the Blue List institutes in east Germany to find a way of ensuring that they have more chance of winning DFG grant money.

According to Ingolf Hertel, director of the Max Born institute for non-linear optics in east Berlin, and spokesman for the Blue List institutes, a possible rule change is now being discussed which would allow scientists at the institutes to apply to the DFG for money for research projects that are in line with the main research interests of their institutes. This cannot be done under current DFG rules. "But it all hangs on the renewal of the five-by-five agreement", says Hertel.

A. A.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Fotostudio Querbach

**Frühwald: optimistic
on future prospects.**

US weapons labs face curb on civilian role

New Mexico. Hundreds of Cooperative Research and Development Agreements (CRADAs), through which national laboratories funded by the US Department of Energy (DOE) have established new partnerships with industry, face dissolution under budget cuts being supported by the Republican majority in Congress.

Proposals now being considered to discontinue funding for 'technology transfer' activities are likely to have a devastating effect on the laboratories. According to Sig Hecker, director of the Los Alamos National Laboratory, collaboration with industry keeps Los Alamos at the forefront of technology; terminating it would damage the quality of all of the laboratory's work.

But Republicans in Congress say that the technology transfer work should end, partly to leave room in the budgets for basic research. "These are nice things to do, but they shouldn't be done at the expense of real scientific work," says a spokesman for Dana Rohrabacher (Republican, California), chair of the energy and environment subcommittee in the House, and an opponent of allocating money for technology transfer.

The question of whether government

should help private industry to develop technology has become a central point of conflict between the Clinton administration and the Republican Congress. The conflict is being watched closely by the vast Los Alamos and Sandia weapons laboratories in New Mexico, which will spend \$150 million between them this year on CRADAs, a sum matched by their industrial partners.

Lacking a development programme for new nuclear weapons, both laboratories are struggling to find a new role — and the development of key technologies of interest to industry has emerged recently as one of their most important activities.

But earlier this year, an independent panel chaired by an industrialist, Bob Galvin, ruled that these activities should not be a "core mission" of the laboratories, sending laboratory managers scurrying to ensure that each CRADA they are engaged in supports their core mission. In the case of Los Alamos and Sandia, that means they must support "stockpile stewardship" — the new and somewhat nebulous role of ensuring the safety and reliability of the US nuclear weapons stockpile (see box).

Sandia is the engineering laboratory

Peace work: Sandia laboratories adapts parachutes developed for nuclear weapons (background) for testing automobiles.

which used to design everything in an American atomic bomb except the nuclear warhead itself. It has a strong tradition of collaboration with industry which it hopes can survive this year's budget upheavals.

Because Sandia is responsible for everything from the springs that hold weapons components in place to the electronics that detonate them, it can sensibly argue that most of its CRADA work is directly relevant to its core mission. For example its work on the dynamics of rubber with Goodyear, a tyre manufacturer, can be justified because the resultant computer codes can be used to update the design of weapons components.

Los Alamos, however, is a scientific laboratory whose special knowledge of fissile materials is less obviously relevant to the outside world, and whose links with industry are newer.

A collaboration with General Motors on the use of plasmas to harden metal surfaces is relevant to the laboratory's work with plutonium components. But officials conceded that other CRADAs which Los Alamos has pursued have had no direct bearing on weapons work.

Al MacLachlan, deputy undersecretary for technology partnerships at the DOE, says that the department is engaged in 1400 CRADAs, and will spend about \$1.5 billion on technology transfer this year. Much of that money is spread across the research in energy supply, which is likely to be virtually halved by Congress.

But the most radical cuts being proposed are in money that is explicitly set aside for the laboratories to spend on technology transfer. \$220 million was allocated this year from the nuclear weapons programme — most of it for Sandia and Los Alamos — and another \$60 million from energy supply research, divided between the DOE's ►

50 years on, success has bitter taste

New Mexico. At 5.30 a.m. on 16 July 1945, the Manhattan project reached its culmination with the successful testing of the world's first atomic bomb at the Trinity test site in southern New Mexico. The plutonium-implosion weapon was tested because its designers were unsure if it would work. A simpler, uranium-based device was used — untested — to attack Hiroshima on 6 August, and a plutonium-implosion weapon used on Nagasaki three days later.

Fifty years on, scientists at the Los Alamos laboratory, where these instruments of mass destruction were conceived, are slowly, and sometimes reluctantly, coming to accept a new regime of no tests, no new weapons, and steady retrenchment in the scope of their work.

At Los Alamos, the number of nuclear weapons scientists has dropped from 1800 eight years ago to 750 today. Officially, this group does not work on new weapons design but on 'stockpile stewardship', relying chiefly on theoretical calculations, test data and computer simulations to assess the future safety and reliability of the weapons stockpile.

But although the weapons team has shrunk, the laboratory continues to employ 7,500 people, spending over

\$1 billion each year on fields ranging from anti-nuclear proliferation to health research. Three-quarters of the work is still related to nuclear weapons in some way, and is paid for either the weapons or nuclear clean-up programmes of the Department of Energy.

Sig Hecker, the director of Los Alamos, concedes that the laboratory's transition since testing stopped in 1992 has been difficult for his staff. But he claims that many of them are now "not only engaged in, but excited by" their new mission.

Jas Mercer-Smith, deputy director of weapons technology at Los Alamos, says assessing the degradation of weapons over time is more technically demanding than their initial design, as they lose their symmetry. "Take a vase, and take a chip out of it," he says. "Has it lost its functionality? This is very challenging science."

But Mercer-Smith and Hecker admit that the laboratory is suffering morale problems as researchers are diverted to non-weapons work — or leave altogether. Mercer-Smith says he is addressing the "demographics problem" on the weapons programme by seeking postdoctorate students to work on related, non-classified projects such as inertial confinement fusion. **C. M.**

MRC rejects call for radiation tests inquiry

London. Britain's Medical Research Council (MRC) is resisting calls for an inquiry into revelations last week by a television documentary that pregnant women and dead children were used during the 1950s and 60s, often without consent, in experiments funded by the council to test the effects of nuclear fallout.

The documentary included interviews with patients — some of whom were never informed of the nature of the tests or, indeed, that they were part of an experiment — who believe that their exposure to radiation may have triggered late onset of cancer.

But the research council has rejected the need for a follow-up inquiry along the lines of one in the United States, where President Clinton has ordered an investigation into military-backed radiation tests on cancer patients between 1940 and 1974 (see *Nature* 367, 303; 1994).

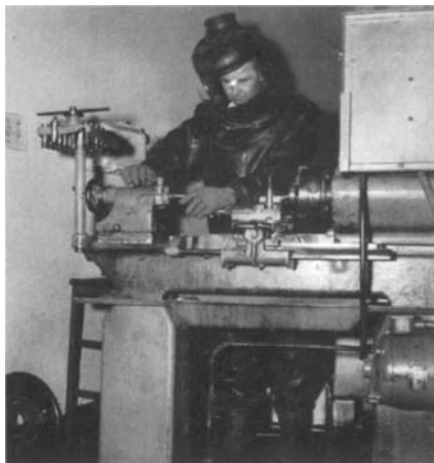
"The MRC will take no action," says David Evered, deputy chief executive of the MRC. He explains that an inquiry would be costly, and that there would be problems in identifying the people involved. "I know this is an important question, but I suggest we have better and more important things to spend MRC money on."

Evered points out that it was "not standard practice to ask permission [for tests on the deceased] in those days", as there was no requirement to seek permission for a *post-mortem* examination before the 1961 Tissue Act. After the programme was transmitted, the MRC set up a telephone helpline for concerned patients and relatives, telling them that there was nothing to fear.

"These were not radiation experiments as

such," says Evered in a statement released after the programme. "Although very small doses of radioactive materials were used in tests in order to help solve important health problems, established techniques were used and considerable research over many years has not shown any adverse health effects from any such tests."

Evered's confidence in the lack of hazards in the experiments however, is not universally shared. Some radiologists contin-



Were patients in radioactivity experiments as carefully protected as workers (above)?

ue to claim that the risk of contracting cancers from exposure to low radiation, though small, should not be ignored.

Others insist the risk needs to be placed in context. "Any radiation dose carries the risk of a later onset of cancer," says one. "But this risk has to be compared with the ordinary risk of contracting cancer, which is

supportive hearings last week, with a view to bipartisan legislation to improve their efficacy. But the Republicans say that CRADAs should compete on equal terms for programme funds.

Officials familiar with the history of the CRADAs, however, doubt that they will be able to do that. The money was set aside after 1989, they say, precisely because the laboratories were failing to establish collaboration with industry using existing programme funds. Industry was interested in such deals, officials say, but given the choice, laboratory scientists preferred to spend their own money on their own projects.

Even 20 years ago it was clear that Sandia and Los Alamos could not survive on nuclear weapons work alone, and since then they have turned for support to the energy programmes launched by President Jimmy Carter, to President Ronald Reagan's Strategic Defence Initiative and, most recently, to technology transfer. The concern of the 16,000 staff at the two laboratories is where can they turn next?

Colin Macilwain

quite high in populations such as ours."

The MRC's decision not to hold an inquiry is likely to remain controversial. Richard Guthrie, a spokesman for the Verification Technology Information Centre (VERTIC) in London, says a thorough investigation would be the best way to assess the extent of radiation damage on those involved in the experiments. "Even if below background level, the radiation dose must have been sufficient to interfere with the body in some way, otherwise there was no point in the experiments," says Guthrie.

The hour-long television documentary, *Deadly Experiments*, based its report on hospital records and research results published in scientific and medical journals. According to the programme's producers John Brownlow and Joe Bullman, both of the London-based production company Twenty-Two Television, the experiments had two broad aims.

One was to evaluate the risk to soldiers in the event of exposure to a nuclear explosion. One set of tests, for example, involved irradiating 6,000 bone specimens taken from dead children in north Wales between 1955 and 1970; another experiment, not linked to the MRC, involved administering total-body radiation to 17 terminally ill patients at the Churchill Hospital in Oxford.

The second aim of experiments was to monitor the effects on the human body of nuclear accidents such as the 1957 fire at the nuclear fuel reprocessing plant at Windscale, in the north of England, which led to a release of radioactive iodine into the atmosphere. The programme-makers allege that in a series of such tests carried out between 1952 and 1957, 30 pregnant mothers were administered radioactive iodine, and the effects later analyzed at the UK Atomic Energy Research Establishment at Harwell in Oxfordshire.

Another MRC experiment involved administering radioactive iron to 21 Punjabi women living in Coventry in the West Midlands. The women were fed chapatis — unleavened bread — laced with radioactive iron, delivered on alternate days.

The women claim they were never informed of the bread's ingredients, or why they were taken to Harwell, where the results were analysed. But the MRC insists that there was nothing secretive about their experiments which, they say, have always been in the public domain and were performed after concern from the department of health of iron-deficiency among Asian women.

The use of radioactive iodine in pregnant women also had a medical purpose, a spokesman says, namely to trace the changing function of the thyroid gland during pregnancy, which often leads to an increase in thyroid diseases.

Ehsan Masood

►non-weapons laboratories. The relevant House committees have reduced these allocations to \$25 million for the weapons laboratories, and zero for the rest.

At Sandia, technology transfer manager Warren Siemens hopes that most of the laboratory's CRADAs — which generally last for three years — will be maintained with money from the weapons stewardship programme.

If there is no extra money for the weapons programme — and the latest House proposal allows none — or if the language says that the CRADAs cannot be rolled over into the main programme, then "we'd have to terminate all the CRADAs and seek to renegotiate them". At that point, he predicts, industry would pack up its bags and go home.

Neither Congressional staff, the DOE or the laboratories themselves are ready to predict exactly what impact the House budget proposal would have on CRADAs. Republicans say they support CRADAs; even as the money for them is being decimated, House subcommittees held

French ethics panel warns of 'crisis' in science reporting

Paris. The reporting of biomedical issues by the media is "a crisis in the making" according to the French national bioethics committee, which last week submitted to the government a statement on the publication of biological and medical information.

The committee says reporting on such issues has become an important social, cultural, economic, political and civic issue. It claims that an "ethical threshold" has now been crossed with the increased occurrence of "rash announcements, retention of information, self-interested connivances, attempts at manipulation of decision-makers, and impenitent spreading of false ideas".

The rapporteurs of the statement — Henri Atlan, professor of biophysics at the faculty of Broussais Hôtel-Dieu in Paris, and the philosopher Lucien Seve — challenge, for example, the coverage given to controversial claims by a French research group that freezing embryos affects their subsequent development, and that the safety of artificial reproductive technologies in humans may have to be reassessed (see *Nature* 373, 553; 1995).

The statement is an extensively revised version of a report on the topic published by the committee last year. In particular, this recommended the appointment of an ombudsman to regulate scientific reporting, warning that if journalists did not address the problems themselves, "the legislator or the government risked imposing measures that would not be very pleasant."

This initial report was vigorously opposed by much of the press and many scientists, who considered it a thinly-veiled attempt to restrict press freedom. The committee has subsequently held discussions with journalists, and taken into account issues raised at a conference organized earlier this year by the French association of science journalists.

The statement still retains the idea of appointing an ombudsman of some sort. But it recommends that he or she should have no legal powers, merely attempting to find solutions to particular problems.

The statement otherwise restates much of the earlier report. Much of the blame, it argues, lies with the media's "obsession with audience figures", its search for "scoops" and its preoccupation with "the sensational and emotional". Medical reporting is criticized as being "too often" a cover for advertising.

But the scientific community also comes under fire. The committee criticizes 'publication by press conference', reiterating that news reports should be based on results that have been published in a peer-reviewed journal. Worldwide media coverage was given, for example, to claims by researchers

at the Pasteur Institute in Paris to have identified the 'door' through which HIV infects cells, whereas the results had not yet been published and indeed were later contested by other groups (see *Nature* 366, 6; 1993).

"Publish or perish" is part of the problem, according to the committee, recommending that research organizations reconsider the means by which they evaluate researchers. Similarly, the committee attacks what it describes as "connivance" between some researchers and journalists, one benefitting from the "scoop" and the other from coverage, and calls for journalists to be given equal access to information.

In this context, the committee criticises *Nature* and other "high-level" journals for "beginning to be affected by the logic of communication", claiming in particular that



by distributing summaries of their articles several days before an issue is published, such journals allow promotional considerations to influence editorial policy.

The committee also draws attention to the risks of disinformation in advertising campaigns run by medical charities, where the need to raise funds may override a desire for accuracy. In particular, the committee says charities should be vigilant not to create false hopes as to when new therapies will become available. Similarly, the statement warns that the growing financial interests in biology and medicine risk compromising the accuracy of information.

But the committee's basic assertion that scientific reporting is deteriorating has still been sharply contested. Speaking at a meeting earlier this year, Philippe Lazar, director general of the national biomedical research organization INSERM, argued, for example, that the quality of reporting is in fact improving, taking into account the "explosion" in both the quantity and complexity of scientific information.

Declan Butler

India's R&D agency is directed to get down to business

New Delhi. 'Making money' is to be the motto of India's leading research agency under its new director-general, Ragunath Mashelkar. The 52-year-old chemical engineer, who took over the reins of the Council of Scientific and Industrial Research (CSIR) last week, says his goal is to make the council's laboratories self-financing "by doing research like business".

The council employs nearly 7,000 scientists in about 40 laboratories specializing in areas that range from drugs and pesticides to petroleum and leather. Research will in future be "market-driven and performance-orientated", with priority given to that which can be patented, says Mashelkar, the youngest scientist to head the CSIR.

Basic research will be supported only if it is novel or related to industry's needs. He promises that CSIR's laboratories will be infused with commercial culture — and will also be given greater autonomy.

Under the new policies, laboratories will be expected to follow a fast track of innovative rather than imitative research, attracting more money from Indian companies and projecting CSIR as what Mashelkar calls "a platform for global industrial research and development". He claims that, if successful, CSIR could become a globally competitive R&D organization comparable to the TNO in the Netherlands.

As part of a general strategy to project a corporate image, each laboratory will set up a "commercial arm" staffed by 10 to 15 marketing professionals. CSIR scientists, meanwhile, will be appointed to the boards of private companies to help them acquire "business sense".

Two of CSIR's chemical laboratories will be restructured along commercial lines, with the expectation that others will follow suit. The agency will also develop a 'blueprint' for attracting expatriate Indian scientists to its laboratories, as well as spearheading a national mission for educating scientists on patents and intellectual property rights.

A new incentive scheme has also been announced, under which scientists developing marketable inventions will be paid 40 per cent of any royalties subsequently received by CSIR. Staff will be encouraged to set up companies to exploit their inventions, and the upper limit — currently \$3,300 — on the amount they can earn through private consultancy work has been abolished.

Through such measures, Mashelkar says he expects the CSIR laboratories to earn at least half their annual budgets through the development and sale of new technologies. At present, their joint earnings are \$50 million — one third of the total budget.

K. S. Jayaraman

Delaney clause heads for the history books

Washington. The Delaney Clause — the controversial amendment in US food safety legislation banning the presence in processed foods of chemicals known to have a carcinogenic effect on animals, however low the risk — appears headed for the chopping block.

The Clinton administration, both Houses of Congress and officials at the Environmental Protection Agency (EPA), responsible for enforcing the act, agree the law needs to be updated, and Congressional action is likely over the next few months. But there is little consensus on what form the changes should take.

Fuelling the reform effort are pesticide manufacturers and food processors, who argue that the clause in the Federal Food, Drug and Cosmetic Act is excessively rigid and unscientific. But some environmental groups remain opposed to any changes, and want stronger enforcement of the clause.

The Delaney Clause, drafted in 1958 by James Delaney, a New York Congressman, bans the presence of carcinogenic pesticides in processed food — the so-called zero-risk standard. But the 1954 section of the food act still allows pesticide residues in raw food, with the agency being required to use risk-benefit analysis to set pesticide tolerance levels that take into account their commercial benefits.

Over time, the EPA found it difficult to enforce the strict standard laid out in Delaney, and in 1987, under the advice of the National Academy of Sciences (NAS), stopped its literal interpretation of the law (although this has remained in force for the equally controversial application of the amendment to food additives).

Instead, it began applying a "negligible risk" standard — defined as causing not more than one additional case of cancer among a million people — when setting tolerance levels of pesticides in both raw foods and processed foods. But the EPA continued its practice of automatically banning a pesticide in raw foods if it had already been outlawed in processed foods.

The academy concluded that modern scientific techniques made it virtually impossible for any processed food to pass the Delaney standard as it is written. It also claimed there was a conflict in the act between the standards applied to raw and processed foods, especially when a pesticide was used on crops headed both for the fresh food market and for processing.

Setting tolerances for pesticide uses has proved slow going for the EPA. The NAS has noted that EPA lacks scientific data on which pesticides are carcinogenic. Pesticides that were placed on the market before EPA began stricter enforcement of Delaney in 1978 have mostly been exempt from review,

even though many are very carcinogenic.

Frustrated over what they claim to be a disregard for human safety, the Natural Resources Defense Council and other environmental groups took the agency to court in California, and won an order forcing it to interpret the law literally and complete a review of every pesticide use based on the law. According to the NRDC, the EPA had

IMAGE UNAMIGABLE FOR COMPLAINT FOR A SHORT REASONS

Killing fields? US pesticide manufacturers are among those protesting against 'zero-risk' regulations.

reviewed only 27 of the 600 active ingredients used in making pesticides.

Under the terms of the court order, the EPA is supposed to revoke pesticides that do not comply with the food safety law, about 65 in total. So far, however, only one such use has been revoked; pesticide manufacturers and food processors have fought revocations in court, and have won nearly every case.

But the California verdict has still caused concern among agrochemical companies and food processors. The National Food Processors Association petitioned the EPA to relax its stance on chemicals in processed foods, and to stop simultaneous bans of pesticides in raw and processed foods.

One agricultural group says that a literal compliance of Delaney has already cost \$400 million nationwide. Jay Vroom, president of the American Crop Protection Association, says that his industry "will be forced to litigate every pesticide".

But NRDC says court action was the only

way to get the agency to act. "EPA's consistent approach throughout the 1980s with respect to carcinogens in food was to ignore or evade that historic statute," Erik Olson, counsel for the group, said last month.

The EPA, meanwhile, has argued that the terms of the order are misguided, as it forces the agency to remove some pesticides that are relatively safe in the quantities they are typically found. Nonetheless, it says that it will continue to enforce Delaney as required by the California courts, and will do so until it is instructed otherwise by Congress. It will decide by December whether or not to end simultaneous bans.

Meanwhile, controversy is raging in Congress over reforming Delaney. The Clinton administration introduced a proposal last year, drawing heavily on the NAS recommendations, that directs the EPA to apply a single standard of negligible risk, including non-cancer risks, to both raw and processed foods. The EPA would have to consider the harmful effects of pesticides on children and pesticide testing and safety checks would be strengthened.

But so far, the only legislator to have picked up the proposals is Henry Waxman (Democrat, California), who introduced a bill into the House of Representatives seeking tough new standards for children.

EPA officials are growing nervous over two alternative proposals, a pesticide reform bill drafted by Thomas Bliley (Republican, Virginia), backed by E. Kika de la Garza (Democrat, Texas) and a broad regulatory-reform bill drafted by Robert Dole in the Senate.

The administration is unhappy with both bills. Lynn Goldman, assistant administrator of EPA's Office of Prevention, Pesticides and Toxic Substances, has said that Clinton will not support the first of these unless it includes stronger enforcement measures and more emphasis on consumer safety. Clinton is not expected to sign the Dole bill as currently written.

Adrienne Appel

Finn tipped for top Brussels research post

Munich. **Jorma Routti**, the president of the Finnish high-technology promotion agency SITRA, is expected to be appointed later this month as the new director-general of the European Commission's research directorate. He will take over from the Italian Paolo Fasella in January.

Routti, who received a degree from the University of California at Berkeley, and remains a university professor, has a reputation in Finland as an efficient administrator with considerable experience in technology transfer.

SITRA's main task is to bridge the gap

between research and industry by obtaining venture capital funding and co-financing a system of cooperative networks between universities and industry.

Routti has also been at the centre of a shake-up in Finland's own research landscape over the past decade (see *Nature* 360, 524; 1994). The country is now one of the highest investors in research among Europe's smaller economies and, partly as a result of SITRA's efforts, the share of its exports based on high technology has risen from 4 per cent in 1980 to 16 per cent last year.

Alison Abbott

MITI clears new path for Japan's universities

Tokyo. As reforms sweep through Japan's higher education system, the bureaucratic walls that have made it difficult for the country's universities to receive government research funds from agencies other than the Ministry of Education, Science and Culture (Monbusho) are crumbling.

This month, the New Energy and Industrial Technology Organization (NEDO), which is affiliated to the Ministry of International Trade and Industry (MITI), expects to receive hundreds of applications from university researchers for 100 large project grants, each worth 50–200 million yen (US\$0.6–2.3 million), to develop the 'seeds' of new technology.

It is the first time that a ministry other than Monbusho has been permitted to offer substantial research funds to universities without interference from the education ministry. The NEDO scheme also marks the beginning of a new era, in which large sums of government money are allocated to universities considered to be carrying out top-rate research.

The NEDO funds are included in a supplementary budget approved by the government in May that was intended to mitigate the effects of the Kobe earthquake, but also includes substantial funds for science.

National research institutes as well as universities are eligible to apply for the project funds, which will be awarded in September after review by a committee drawn from universities, national institutes and the private sector. Research fields to be covered include biotechnology, new materials, electronics and information technology, medical, energy and environmental technology.

NEDO's "epoch-making" scheme came as a "surprise", says Hiroshi Saito, a professor at the Tsukuba Advanced Research Alliance (TARA). This is a new organization set up last year by Tsukuba University to encourage joint research between the university, national research institutes and private companies in Tsukuba science city

north-west of Tokyo (*Nature* 369, 268; 1994).

Saito says the project, as well as other related government policy changes, marks an "excellent opportunity for universities to make themselves more open to society and



'Japan: time for change': Tokyo University, from special issue of *Nature*, October 1992.

to expand research by getting a lot of support from outside".

NEDO is not the first non-Monbusho organization to provide large research funds to university researchers. But it is the first to secure the support of the education ministry, which has up to now maintained a bureaucratic stranglehold on university research.

The Science and Technology Agency (STA) has for many years run the Exploratory Research for Advanced Technology (ERATO) programme that provides large project funds to teams of researchers, often headed by university professors.

But such project leaders have to set up an off-campus project office, and cannot officially carry out ERATO research in the university. Indeed, it is widely rumoured that university researchers who accept ERATO funds risk being cut off from funding from Monbusho.

Hiroyuki Kishi, deputy director of MITI's Office for Industrial Science and Technology Frontier Programme in the Agency of

Industrial, Science and Technology, which oversees the new NEDO scheme, says MITI has had frequent contact with Monbusho over the new programme, and that "no special problems" have arisen.

Kishi says that university researchers who win project funding will be free to buy equipment and to use existing university facilities. They will even be able to hire post-doctoral fellows through NEDO.

The NEDO scheme is one of a number of new programmes through which large sums of government money are being allocated to specific universities. This contrasts with Monbusho's traditional approach of providing very small amounts of research funding to all university teaching staff, regardless of what research — if any — they do.

In the current fiscal year the education ministry will provide about 6 billion yen (\$70 million) of extra money to selected university-related institutes, as well as a handful of chosen teams of university researchers, in an effort to create centres of excellence (see *Nature* 371, 188; 1994).

Monbusho has also won 2 billion yen in the supplementary budget for a scheme to be run by the Japan Society for Promotion of Science. This will provide 20 project grants of comparable size to NEDO's to support joint university/industry research.

Saito of TARA says that such moves demonstrate that Monbusho "has begun to endeavour to return the research results of universities to industry". The days of the "ivory tower", he suggests, are now over.

But sceptics say much will depend on how university researchers are selected for funding. In the past, a few powerful professors, with strong links to Monbusho, have tended to choose where the largest grants go.

Critics, including those awarded large Monbusho grants, argue that only with a truly open system of competitive peer review — perhaps involving foreign researchers — will a real revolution take place in Japan's university research. **David Swinbanks**

Report claims that United States is producing too many PhDs

San Francisco. US universities are awarding 25 per cent more doctorates in science and engineering than the economy can absorb, according to researchers at the Rand Corporation and Stanford University's Institute for Higher Education.

"There are institutions producing doctorates whose decision to stop doing so would probably not be noticed by the world," says William Massy, director of the institute, and principal investigator in a new study of postgraduate training. He says that in fields where those universities are not strong, they ought to go out of the business.

Together with Charles Goldman of Rand, Massy analysed 13 science and engineering fields at 210 universities, as well as at 1,000 educational institutions that employ doctoral graduates. They found that, on average, 200 new PhDs in each field fail to find employment related to their degree qualifications every year.

Massy says admission policies to doctoral courses have little relation to the potential job market, but correlate more closely with a particular department's needs for teaching and research assistants.

The report follows on the heels of an analysis by the National Academy of

Sciences that came up with similar results, but drew very different conclusions. The NAS report found more than half of new PhD graduates find work in non-academic settings. But its authors suggested that employment outside academic institutions does not necessarily represent a misuse of the doctoral programme.

Massy disagrees. He says society needs to take a hard look at the cost-benefit equation of employing doctorates in jobs that do not require such rigorous training. "When you consider the social cost of training a PhD, I don't think it pays out," he says. **Sally Lehrman**

Einstein and Lorentz

SIR — Michaël W. J. van den Brink recalls a sombre occasion when Einstein acknowledged his profound scientific debt to H. A. Lorentz¹.

When, in 1920, Lorentz snubbed Max Planck's sixtieth birthday celebration in solidarity with ascendant post-1918 anti-German revanchism, he was rebuked by Max Born for his blindness to Lorentz's primitivism². Indeed, in 1944, Einstein echoed Lorentz's absolute anti-Germanism when he wrote: "The Germans as an entire people are responsible for these mass murders and must be punished as a people..."³ Einstein still felt the same way in 1953; again, the usually deferential Born was compelled to remonstrate in a reply to a letter from Einstein: "I only want to tell you that the German Quakers have their headquarters in Pymont. They are no 'mass-murderers', and many of our friends there suffered far worse things under the Nazis than you or I. One should be chary of applying epithets of this sort. The Americans have demonstrated in Dresden, Hiroshima and Nagasaki that in sheer speed of extermination they surpass even the Nazis."⁴

But Einstein's grandest and purest genuflection to Lorentz came in 1954: "In January 1954 Einstein was asked: Who were the greatest men, the most powerful thinkers whom he had ever known? He answered without hesitation, 'Lorentz'." Then he added: "I never met Willard Gibbs; perhaps, had I done so, I might have placed him beside Lorentz."⁴

William Steinsmith

1239 Castenada Avenue,
San Francisco,
California 94116, USA

SIR — Einstein had acknowledged his debt to Maxwell and Lorentz unequivocally in his address of 15 May 1940 to the 8th American Scientific Congress in Washington, DC. An abridged version was published in *Nature* a month later⁵. Einstein said:

The theory of relativity arose out of efforts to improve, with reference to logical economy, the foundation of physics as it existed at the turn of the century. The so-called special or restricted relativity theory is based on the fact that Maxwell's equations (and thus the law of propagation of light in empty space) are converted into equations of the same form, when they undergo Lorentz transformation. This formal property of the Maxwell equations is supplemented by our fairly secure empirical knowledge that the laws of physics are the same with respect to all inertial systems. This leads to the result that the Lorentz transformation — applied to space and time coordinates — must govern the transition from one inertial system to any other. The content of the restricted relativity theory can accordingly be summarized in one

sentence: all natural laws must be so conditioned that they are co-variant with respect to Lorentz transformation.

Einstein was not without faults, but non-attribution of credit to Maxwell and Lorentz for his ideas on relativity is not one of them.

Sachi Sri Kantha

5-16-305 Tsukimicho,
Fukuroi City,
Shizuoka 437-01,
Japan

1. van den Brink, M. W. J. *Nature* **375**, 272 (1995).
2. *Born-Einstein Letters* (Macmillan, London, 1971).
3. Einstein, A. *Ideas and Opinions*, 212.
4. Mehra, J. *Physica* **79A**, 474 (1995).
5. *Nature* **145**, 920-924 (1940).

Lucifer principle

SIR — In *The Lucifer Principle*, Howard Bloom attempts to write a world history from a modern scientific viewpoint, but your reviewer seems to have mistaken it for an evolutionary tract whose main purpose is to defend group selection against individual selection. Only a few paragraphs were in fact devoted to that topic, and the main part of the book is about the past few thousand years of human history. For this, it does not matter whether group selection is a common occurrence or a very rare one.

What Bloom draws attention to is the firm grip that human culture holds on the individual human mind, and the role this relationship has played in the horrors and triumphs of human history. Human social culture is radically different from that of other animals because it develops by acquiring new knowledge and incorporating it for transmission to succeeding generations; this process is of course orders of magnitude faster than Darwinian evolution. The development of human culture, and the human individual's almost total (though seldom self-acknowledged) submission to it, have been hugely important in giving us a dominant position on this planet, where we pose a serious threat to many of its other inhabitants. Though it is only a single episode, and no one knows how it will end, does not the human population explosion constitute a proof that group selection can occur?

Your reviewer (Mark Pagel, in *Nature* **374**, 828; 1995) is a member of the BBSRC-NERC Ecology and Behaviour Group in Oxford, which must be vitally concerned with the reasons for humanity's competitive success; it is therefore curious that he should dismiss with a genteel cry of "déjà vu" the main part of Bloom's message.

Horace Barlow

Trinity College,
Cambridge CB2 3EG, UK

Free exchange

SIR — In response to your article on the French genome programme (*Nature* **375**, 175; 1995), we should like to point out, on behalf of the members of the Washington University Genome Sequencing Center and the Sanger Centre, that our respective laboratories have always freely and promptly released their data without restrictions of any kind, and will continue to do so in the future. As far as we are concerned, there is therefore no question of France being in a "position of dependence" as the French report you refer to claims.

Furthermore, we have explicitly put forward the human sequencing project as a subject for international cooperation (see, for example, *Nature* **375**, 93; 1995). We intend to do our bit, but the project can be completed only if others will do theirs. With the map coming to fruition, the human genome can readily be divided by chromosome or by region for the sequencing task, just as the yeast genome was successfully divided for the same purpose. If all the national programmes contribute, the community can have the human genome sequence in hand by the year 2001.

Bob Waterston

John Sulston

The Sanger Centre, Hinxton Hall, Hinxton,
Cambridge CB10 1RQ, UK

Remarkable delay

SIR — I was surprised to read the "Corrigendum" in *Nature* (**375**, 88; 1995) to a paper published many months ago (T. D. White *et al.* **371**, 306-312; 1994). Even if the correction was to remedy a serious typographical error or other inadvertent mistake, the delay would be remarkable; but it is clear from the Corrigendum that the error is not inadvertent as it seeks to create a new genus in the Superfamily Hominoidea on the grounds that new finds made subsequent to the initial publication render the original conclusions unacceptable to the authors. The new finds are alluded to and a diagnosis is given in the Corrigendum, but the material is neither described nor illustrated.

It is, of course, the right of the authors to alter their views on the classification of their materials, but it is also their duty to make available the evidence for these changes. This duty will no doubt be fulfilled in due course. There may be some, however, who will feel that the hasty erection of a new genus in this way has more to do with expediency and priority than with scientific enlightenment. Others may feel that the editor has been bowled a fast straight one that took out his middle stump!

M. H. Day

Natural History Museum,
Cromwell Road, London SW7 5BD, UK

Aquinas and the embryonic life

SIR — My hope, in writing about the Pope and human development¹, was to build on some existing Roman Catholic thinking on the basis for humane exceptions², and on Aquinas and reproductive issues, because of their moral and practical importance. If persuaded that this would be right, John Paul II has the bravery and force of character to do great practical good for people in general. I am sorry that my argument caused offence to a number of your readers.

I was encouraged to write by the infectious enthusiasm of St Thomas Aquinas to understand the natural world, for which he relied on Aristotle. When, in 1261, he read the new translation, by William of Moerbeke, of Aristotle's treatises on animals, he urged William to continue translating. As further books came to him in accurate Latin, he would dictate his own work, like a chess grand master, to several secretaries simultaneously. Aquinas argued that the soul is logically related to the body, and ensoulment is gradual³, dependent upon the development of the organs necessary for intellectual and spiritual life⁴.

Only a month ago, Pope John Paul II wrote an encyclical letter, *Ut Unum Sint*, seeking reunion of the Catholic and Orthodox churches. So Aquinas deserves particular attention just now: in 1274 he was summoned to a general council in Lyons on reconciliation of the Christian Churches, and he died on his way to it. Aquinas could yet inspire appropriate guidance on the ethical basis for the family and the population in today's world, so allowing churches an ecumenical future, as he earlier foreshadowed the renaissance.

John Godfrey

41 Lawford Road,
London NW5 2LG, UK

1. Godfrey, *J. Nature* **373**, 100 (1995).

2. O'Donoghue, N. D. *Irish theolog. Q.* **35**, 217–232 (1968).

3. Aquinas, in *111 Sententiarum*, dist.3, q.5, art.2; and *De anima* art. 11.

4. Donceel, J. *Theolog. Stud.* **31**, 76–83 (1970).

● This correspondence is now closed. — Editor, *Nature*.

Politically correct

SIR — In commenting on the resignation of Martin Harwit from the National Air and Space Museum in Washington, DC (*Nature* **375**, 95; 1995), Stanley Goldberg is right to say that scholars should be able to "do their work without having to make sure that it agrees with prejudice of outsiders".

It would also be desirable, however, if the same scholars were not so pressured from insiders either. Everybody knows that in academic institutions (especially in non-science subjects), the shortest and safest way to promotion and tenure and to key committees is political correctness. This includes blaming the Allies in the Second World War for using the atomic bomb or for bombing Dresden. (The logical corollary is that the Allies should not have been in such a hurry to end a war that was devastating to both sides.)

Being politically correct is also very useful outside academic life, especially if one wants to be interviewed by public radio or to be cited by, say, *The New York Times*. The pressure to conform is as strong from the inside as from the outside, which leads today's scholars to use collective minds instead of their own individual minds.

I believe that there was no need for the resignation of Martin Harwit; all that was required was that he (and other administrators following in his footsteps) should keep in mind that, in controversial matters, a scholar is supposed to present both points of view. Anyone who lived through the Second World War remembers how much hatred was abroad in those days everywhere. And if it is true that the Allies sometimes reacted in anger, it is also true that they had good reason to be angry.

Renato Baserga

125 Bledwyn Road,
Ardmore,
Pennsylvania 19003, USA

Friedmann models

SIR — John Maddox considers it striking that Tolman's book *Relativity, Thermodynamics and Cosmology* did not mention Friedmann's world models (*Nature* **375**, 445; 1995). He believes "they languished in the Russian language until about the publication of Tolman's book". In fact Friedmann's famous cosmological papers were published in German in the *Zeitschrift für Physik* (**10**, 377; 1922 & **21**, 236; 1924) which was a widely read journal at that time. This makes Tolman's ignorance (and that of others) even more striking.

Friedmann discussed the solutions in terms of elliptical integrals and retained the cosmological constant λ . He did not give the more explicit simple solutions for the special case $\lambda = 0$ which are often erroneously attributed to him.

It may also be of interest that the translator of a recent Russian paper (Frenkel, V. Ya. *Physics-Uspekhi* **37**, 767; 1994) states that the correct transcription of Friedmann's name should be A. Fridman. As Friedmann's papers are written in excellent German, one wonders which

transcription he preferred. In his first paper (to which he refers in the second as "unserer Notiz", or "our note") the author's name is A. Friedman, in the second A. Friedmann.

Hartmut Schulz

Astronomisches Institut,
Ruhr University,
Bochum,
Germany

One 'f' in trophic

SIR — John Funder is most concerned about the second 'p' in apoptosis (*Nature* **371**, 98; 1994 & **373**, 379; 1995) and has fired considerable interest (*Nature* **372**, 312; 1994 & **374**, 670; 1995). But as an endocrinologist I suggest that the p in trophic should also concern us.

Over the past few years, this letter has become an endangered species. Led by the Americans, it has been hunted almost to extinction. Wonderful hormones such as adrenocorticotrophin, thyrotrophin and somatotrophin have become but inaccurate shadows of their former selves.

Surely, the stem must come from τροφικῆς ('nourishment', which is what each of these hormones gives to its target cells) and not τροπικῆς ('turn', as in geotropism)?

In this case it is not pronunciation that matters but accuracy.

Stephen Nussey

Division of Biochemical Medicine,
St George's Hospital
Medical School,
Cranmer Terrace,
London SW17 0RE, UK

Mayflies in May

SIR — As a fisherman, I am interested in the hatching of the mayfly, *Ephemera danica*. This happens over about two and half hours in the early evening and continues for about a fortnight. It is known traditionally as 'Duffers' fortnight', as even the most incompetent dry fly fisherman may catch a fish or two.

This year the hatch here, on the Gloucestershire Coln, was roughly between 21 and 28 May. The point of scientific interest is that a local sportsman fishing the same river at the turn of the century wrote that the keeper reckoned that the best hatch was on Derby day (the first Wednesday in June)¹. This means that the hatch today is happening some ten days earlier than it did a hundred years ago.

W. B. Stawell

Taylors Farmhouse,
Aldsworth,
Cheltenham GL54 3QX, UK

1. Gibbs, J. A. *A Cotswold Village* p.154 (1898).

Restoring good manners in research

The need for good manners arises from a conflict in the foundations of the research enterprise. Journals have much to do to put things right, but the chief responsibility rests with academic institutions.*

Is there an intrinsic conflict in the principles on which the scientific enterprise is based? Everybody agrees that science is a cumulative enterprise in which people stand on others' shoulders "to see further"; Copernicus and Galileo were among "the giants" on whose shoulders Newton stood, but there were Galileo's lens grinders and polishers whom he did not mention. Now, the cumulative character of the enterprise means that spectacular discoveries (say the structure of DNA) are rare, but that lesser discoveries are necessarily precursors and are therefore equally estimable. Good science is good science, whatever its importance, and its practitioners are equally deserving of respect. International interdependence flows from that.

Second, publication is of the essence. A discovery has no meaning unless an account of it is generally available. Paradoxically, equipment is more vulnerable to the passage of time than an account in words of the use made of it. But the function of publication is more than the transmission of the news to others. It is the only means by which the record of discovery can be authenticated in the long run; critical readers pick holes in the logic, others test for consistency by telling whether some field can be welded into a coherent whole. (One of the problems yet to be solved in electronic publications is that of telling which is the authentic text.)

Third, the traditional and, until recently, the sole reward of the working scientist has been his or her bibliography. It is a curiously insubstantial proof that a lifetime's science has been well done, but when the deepening of understanding is a cumulative business, even a handful of good papers rightly engender pride. The other traditional component of the profession's reward structure is the joy of teaching students successfully. It is impressive how often senior people list the names of erstwhile graduate students who are themselves now senior and productive people. The solicitude of the profession for its students is its most appealing attribute.

Where, in this account of an altruistic research profession, can there be conflict? If publication is essential to the authentication of discovery and also relevant to a researcher's self-esteem (and to the esteem accorded him or her by others), how can (and how strongly should) the temptation to improve the lustre of a personal bibliography be resisted? Succumbing to temptation distorts, even corrupts, the record.

That is why there had emerged in the old days (and people will differ over how long ago they were) a code of conduct, an etiquette, to which people adhered more or less scrupulously.

Good manners were simple. A person should write and talk openly about past accomplishments and even future plans, answering intelligent questions even from competitors. Research data and materials should be shared with serious people, at least after the relevant papers had been published. Peripheral contributions to a piece of work should be generously acknowledged for what they are, and not dignified by co-authorship. Although Sir William Ramsay's plea a century ago that people should not follow up a discovery without the permission of the discoverer asks too much of science as well as its practitioners, to fail to give full credit to pathfinders in a field is not merely discourteous, but a way of robbing them of their own place in the record of discovery. And people established in a field have a professional duty to comment on the work of others in the capacity of referees.

Fifty years ago, these principles were widely followed, even in busy and productive university departments. The principles are still understood and still widely followed. Journals are especially conscious of what this entails. *Nature's* demands on its referees are clamant and onerous; it is a constant source of wonder that so many offer such careful, judicious and constructive reports on manuscripts people working in fields related to their own interests are anxious to have published. Good manners still obtain.

But there is evidence in the hands of all journals that bad manners increasingly coexist with good. Self-advertisement is more common now than it used to be. People refer to their own publications when others' would be more apt. When reference to competitors is unavoidable, they may refer to a minor paper, ignoring a more important work. (But not so long ago, a Japanese author who will recall the circumstances referred to an inappropriate publication by a prominent US scientist in the belief that such modesty would help ensure publication.) The tendency to overlook publications in other languages, or even by colleagues in other countries, is a growing source of angst. So to is the use of the word "first", as in "We have for the first time shown that Ohm's Law is valid in spacecraft beyond Jupiter (or Neptune, or

Pluto, as the case may be).

The perils of honorary co-authorship are now well documented, not least by Feder and Stewart's investigation of the circumstances in which John C. Darsee, at Emory University and the Harvard Medical School, was routinely able to recruit distinguished people as fellow-authors. The difficulty is that people with responsibility for administration and fund-raising are supposed to be more effective at those tasks if they can also boast of a research record.

There are also lapses from good manners by the gallant company of referees. As long ago as 1967, a paper on the metallurgy of austenitic steel was returned from a referee with the opinion that it did not deserve publication; the referee enclosed a paper of his own saying much the same thing, not manifestly more compellingly. Cases in which the sight of an authors paper has stimulated referees to publish their own parallel work elsewhere have been recorded. The assumption that referees always treat a manuscript as confidential seems not always to be valid.

What can be done to bring back good manners? Journals have a responsibility to be more severe on offending manuscripts, and to make public flagrant transgressions of the unwritten rules. They have a particular duty to fight against that form of obscurity that stems not from a poor grasp of language, but from a wish to shelter from clarity in case it should be more vulnerable to criticism. It may make sense to ask corresponding authors to vouch for the collective responsibility of all co-authors.

But academic institutions and grant-making agencies have the chief responsibility. The temptation to magnify an individual bibliography, always present and usually resisted, is magnified by the pressure to win promotion and research funds, where decisions may hang on bibliometric indices of some kind. Because the importance of a discovery may seem, at least in its authors' minds, to be magnified by reports in the general press and thus by being first, anxiety over priority mounts.

Yet little thought has been given to mechanisms for decreasing the pressure to publish and to be first — especially in the United States where they are strongest, and which sets the tone for the rest of science. Until that is done, good manners will remain under threat. **John Maddox**

* Gist of an address to the International Federation of Science Editors, Barcelona, 9 July.

Changes in the Zodiacal Cloud

George J. Flynn

AT present about 4×10^7 kg — 40,000 tonnes — of interplanetary dust from the Zodiacal Cloud falls onto the Earth annually¹. But has this rate been constant through time? Apparently not, according to evidence presented by K. A. Farley on page 153 of this issue².

Farley has measured the concentration of ³He solar ions implanted into interplanetary dust, in an oceanic core spanning the period from 0.19 to 69.27 million years ago. After correction for the varying sedimentation rate, Farley finds two maxima in the ³He concentration, occurring 37 and 50 million years ago. He interprets these maxima as periods when dust accretion was a factor of three higher than the mean over the interval from 10 to 20 million years ago, and a factor of six higher than the minimum value (16 million years ago). These results clearly call into question any assumption that, over the past few billion years, the Zodiacal Cloud has been in a steady state, producing a constant interplanetary dust flux at Earth.

Small interplanetary dust particles spiral rapidly into the Sun because of the Poynting–Robertson effect, a process of orbital decay caused by drag from solar radiation. Larger particles become smaller by erosion or catastrophic collisions. It takes under 100,000 years for a typical particle of 10–20 micrometres in diameter, starting at 3 astronomical units out, to spiral into the Sun. So the interplanetary dust presently accreted by the Earth was emitted from much larger objects in the relatively recent past. A constant source of dust, providing about 9×10^3 kg s⁻¹, is required to maintain the Zodiacal Cloud in a steady state³.

Until recently, comets were believed to be the main source of resupply. But Kresak⁴ showed that the total mass of dust emitted by currently active comets provides only 2 per cent of the necessary amount. Whipple⁵ proposed that the present Zodiacal Cloud was generated by comet Encke during an early and much more active phase of its evolution, an implicit acceptance that the Zodiacal Cloud is not in steady state. The discovery of dust bands associated with asteroid families⁶, and physical evidence from interplanetary dust collected from the Earth's stratosphere⁷, suggested that some of the dust comes from asteroids. But analysis of the dust-band data by Dermott *et al.*⁸ indicates that main-belt asteroids contribute only 30 per cent at most. In all, the inadequacy of identified dust sources to maintain the Zodiacal Cloud in steady state supports Farley's observation of a time variation in the

accretion rate, and further suggests that we are currently in an era of increased dust flux.

This has implications for the sources of the dust. The slowly eroding surfaces of all asteroids and active comets contribute continuously to the interplanetary dust. But from the large peaks in the dust flux, as reported by Farley, it would seem that one, or only a few, dramatic events, such as the catastrophic disruption of a large asteroid or the appearance of an unusually active comet, contribute the bulk of the dust in times of peak intensity.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

'Zodiacal light' seen after sunset as sunlight reflects off zodiacal dust.

During these periods of higher flux, most of the interplanetary dust should reflect the composition and mineralogy of one or a few parent bodies. The abundance of carbon, which is particularly diagnostic of meteorite type and albedo, has been measured in some 60 stratospheric interplanetary dust particles^{9,10}. All have carbon contents greater than the carbonaceous chondrite Allende^{9,10}, suggesting that they do not span the range of albedos, and inferred carbon contents, exhibited by main-belt asteroids^{11,12}. Even the most diverse types of interplanetary dust, the hydrated and the anhydrous particles, may originate from a single parent body¹³. This lack of compositional diversity is consistent with the bulk of the interplanetary dust presently falling onto the Earth being derived from one or a few dramatic events^{11,12}.

An alternative to Farley's interpretation is that these variations record changes in the solar emission rate of ³He, resulting in higher ³He concentrations in dust particles exposed during peaks in the emission. Examination of lunar samples and meteorites shows no evidence that there is much variation in the solar ion flux,

although these measurements are not particularly sensitive to short-term variations of a factor of three or so. So the cause of the ³He spikes at 37 and 50 million years ago must be investigated by other techniques.

If, as Farley proposes, ³He has not diffused from its original site and the record is therefore not distorted, intact interplanetary dust from earlier times should be preserved in the core. Farley calculates a concentration of about 7 parts per million of small (up to 50 micrometre) interplanetary dust particles, and up to 700 p.p.m. of larger particles in the sediment that is richest in ³He. These larger particles should be identifiable by mapping sediment sections for elements rich in chondritic particles, followed by chemical and mineralogical studies on the chondritic regions. If Farley is correct, comparison of the 16- and 37-million-year-old layers will show dust concentrations that vary by a factor of six. Furthermore, in the layers recording peaks in the dust flux, the composition and mineralogy of the dust should reflect that of the parent bodies causing the enhancement.

Farley points out that iridium, which is normally used as a tracer of extraterrestrial matter, does not correlate with ³He. This is because iridium traces the total deposition of extraterrestrial material, whereas ³He traces only those extraterrestrial particles that are small enough (less than 50 micrometres) to accumulate a significant bulk concentration of solar ³He and retain that ³He during entry into Earth's atmosphere. Particles smaller than 50 micrometres constitute about 10 per cent of the total mass of interplanetary dust¹, so Farley's results relate directly to only a small fraction of the interplanetary dust falling onto the Earth. Direct identification of dust particles in the sediments would track the flux of particles over the whole size range. □

George J. Flynn is in the Department of Physics, State University of New York at Plattsburgh, Plattsburgh, New York 12901, USA.

1. Love, S. G. & Brownlee, D. E. *Science* **262**, 550–553 (1993).

2. Farley, K. A. *Nature* **376**, 153–156 (1995).

3. Grün, E. *et al.* *Icarus* **62**, 244–272 (1985).

4. Kresak, L. in *Solid Particles in the Solar System*, IAU Symp. 90 (eds Halliday, I. & McIntosh, J.) 211–222 (Reidel, Dordrecht, 1980).

5. Whipple, F. L. in *Zodiacal Light and the Interplanetary Medium* NASA SP-150 (ed. Weinberg, J. L.) 409–426 (NASA Sci. Tech. Information Office, Washington DC, 1967).

6. Low, F. J. *et al.* *Astrophys. J.* **278**, L19–L22 (1984).

7. Flynn, G. J. *Icarus* **77**, 287–310 (1989).

8. Dermott, S. F. *et al.* in *Asteroids, Comets, and Meteors 1991* (eds Harris, A. W. & Bowell, E.) 153–156 (LPI, Houston, 1992).

9. Schramm, L. S. *et al.* *Meteoritics* **24**, 99–112 (1989).

10. Thomas, K. L. *et al.* in *Analysis of Interplanetary Dust AIP Conf. Proc.* 310 (eds Zolensky, M. E. *et al.*) 165–172 (American Institute of Physics, New York, 1994).

11. Flynn, G. J. *Meteoritics* **28**, 349 (1993).

12. Flynn, G. J. *Planet. Space Sci.* **42**, 1151–1161 (1994).

13. Zolensky, M. E. & Barrett, R. *Microbeam Anal.* **2**, 191–197 (1993).

Back to primary school

John P. Moore

ATTACK by the human immunodeficiency virus on T cells involves binding of the viral envelope gp (for glycoprotein) 120 subunit to the cells' CD4 receptor. Binding is followed by fusion of virus and cell, the first step in the viral replication cycle. Because gp120 is central to this process, using gp120 subunit vaccines to generate neutralizing antibodies against HIV-1 has been at the forefront of vaccine strategies for many years. Yet there has been vigorous debate about whether such an approach will work.

That debate now takes on a new dimension with the appearance of a paper by Sullivan *et al.* in *Journal of Virology*¹. In general, the findings bear on the issue of the relative resistance of primary strains of HIV-1, as opposed to strains highly passaged *in vitro*, to neutralizing antibodies, and whether such variable resistance is real or an assay-dependent artefact. In particular, the new results may help explain why antibodies raised to current monomeric gp120 vaccines do not neutralize primary viruses^{2,3}.

Neutralization

Sullivan *et al.* have measured the interactions of monoclonal antibodies (mAbs) with the native, oligomeric forms of the HIV-1 envelope glycoproteins, found on virions and virus-infected cells, and compared the results with levels of virus neutralization. They draw various conclusions. First, they confirm that mAb interactions with monomeric gp120 do not predict the extent of virus neutralization⁴. Second, they take things further and show that mAb reactivity with oligomeric glycoprotein is a good correlate of neutralization, which is consistent with observations on HIV-1 LAI from Sattentau and myself⁵.

The third conclusion is that mAbs don't bind nearly as well to oligomers from primary viruses as to oligomers from strains whose envelope glycoproteins have been altered by passage in T-cell lines *in vitro*^{1,6,7}. The relative inaccessibility of neutralization sites on primary viruses, including the V3 loop, combined with Sullivan and colleagues' observation that such viruses have a greater density of surface glycoproteins, may account for much of the mechanism behind neutralization resistance. HIV-1 has presumably evolved these defences to counter the abundant antibodies raised against it in infected humans, perhaps at the expense of some loss of replication efficiency^{1,4}. As disease progresses and the immune system collapses, the more virulent strains that can emerge may need less protection.

Resistance of primary virus to neutralization

was first documented from studies on soluble CD4 (sCD4)⁸, and is known to be a function of the oligomeric structure of the envelope glycoproteins⁹. But it has not been widely appreciated that resistance to neutralization by antibodies probably has the same basis, in that sCD4 resistance is an indirect consequence of a more natural phenomenon. Soluble CD4 strongly neutralizes strains of HIV-1 adapted to T-cell lines; yet it increases the infectivity of certain strains of simian immunodeficiency virus and HIV-2¹⁰, which led to the proposal that SIV's capacity for fusing with a cell ('fusogenicity') is activated by binding of CD4¹¹.

It now seems clear that HIV shares this property. Sullivan *et al.* show that sCD4 has a biphasic effect on the fusogenicity of two strains of primary HIV-1, increasing it at low concentrations, but inhibiting it at higher concentrations as gp120-sCD4 complexes are stripped from the virus. Schutten *et al.*¹² have also observed sCD4-enhanced fusogenicity of primary HIV-1 clones. Previous studies with sCD4 helped create the perception that the envelope glycoproteins of HIV-1 were in some critical way different to those of other primate lentiviruses. But it now seems clear that the anomalous viruses are the highly adapted strains of HIV-1 such as LAI and MN, not HIV-1 *per se* — a lesson we are learning, once again, the hard way.

Yet perhaps even more striking are results stemming from Sullivan and colleagues' experiments with a human mAb (F105) to a complex gp120 epitope near the CD4-receptor binding site. F105, they find, not only fails to neutralize a primary strain of HIV-1, but actually increases its infectivity. It does, however, effectively neutralize strains that are adapted to T-cell lines.

These observations are akin to those with sCD4 — by binding to an epitope overlapping the CD4-binding site, an antibody seems to be able to trigger some of the fusion-activating conformational rearrangements in the viral envelope. As different mAbs to this epitope cluster induce different conformational changes in gp120 monomers, whether an antibody mediates activation of fusion^{1,12}, or neutralization^{1,4}, might be a subtle function of the geometry of its epitope in relation to the CD4-binding site. Schutten *et al.*¹² have also shown that two other human mAbs to CD4-binding-site epitopes, and one to the V3 loop, increase the fusogenicity of a primary HIV-1 clone; the latter finding is in itself notable, for there is a growing understanding that the V3 loop is not the principal neutralizing determinant for primary viruses.

What are the implications of all this for vaccine development? Several high-quality, monomeric gp120 vaccines have been tested in humans. One advantage of the correctly folded forms of these proteins has been their ability to induce antibodies to the complex epitopes surrounding the CD4 binding site; indeed rodent mAbs have been raised against gp120 that are indistinguishable from antibodies made in infected humans¹³. Yet although sera from gp120-immunized animals and humans neutralize viruses adapted to T-cell lines, they are inactive against primary strains of HIV-1^{2,3}.

Infectivity

There is, as yet, no evidence that the increase in infectivity mediated by F105-like antibodies contributes to such lack of activity, but the possibility should now be considered. We have found that many sera from HIV-1 infected humans, selected for high titres of anti-gp120 antibodies, weakly increase the infectivity of primary HIV-1 strains; other sera merely lack net neutralizing activity (Y. Cao *et al.*, unpublished data). Perhaps F105-like antibodies contribute to this effect by counteracting, wholly or in part, any neutralizing antibodies present (although the involvement of antibodies that increase infectivity in more conventional ways¹⁴ should not be discounted).

The present generation of monomeric gp120 immunogens seems unable to breach HIV-1's defences, and Sullivan *et al.* show just how formidable is the wall that lies across the path to an antibody-mediated, anti-HIV vaccine. We now must decide whether to blast a way through this barrier, subtly undermine it or even choose another route entirely. But brute force is rarely the answer, so the need for a large and costly infrastructure to evaluate the available gp120 vaccines should be reconsidered. □

John P. Moore is at The Aaron Diamond AIDS Research Center, 7th Floor, 455 1st Avenue, New York, New York 10016, USA.

- Sullivan, N., Sun, Y., Li, J., Hofmann, W. & Sodroski, J. *J. Virol.* **69**, 4413–4422 (1995).
- Matthews, T. J. *AIDS Res. hum. Retrovir.* **10**, 631–632 (1994).
- Wrin, T. & Nunberg, J. H. *AIDS* **8**, 1622–1623 (1994).
- Moore, J. P. *et al. J. Virol.* **69**, 101–109 (1995).
- Sattentau, Q. J. & Moore, J. P. *J. exp. Med.* **182**, 185–196 (1995).
- Bou-Habib, D. C. *et al. J. Virol.* **68**, 6006–6013 (1994).
- Stamatatos, L. & Cheng-Mayer, C. *J. Virol.* (in the press).
- Daar, E., Li, X. L., Moudgil, T. & Ho, D. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6574–6578 (1990).
- Moore, J. P., McKeating, J. A., Huang, Y., Ashkenazi, A. & Ho, D. D. *J. Virol.* **66**, 235–243 (1992).
- Allan, J., Strauss, J. & Buck, D. *Science* **247**, 1084–1088 (1990).
- Allan, J. *et al. Science* **252**, 1322–1323 (1991).
- Schutten, M., Andeweg, A. C., Bosch, M. L. & Osterhaus, A. D. M. E. *Scand. J. Immun.* **41**, 18–22 (1995).
- Nakamura, G. R. *et al. AIDS Res. hum. Retrovir.* **8**, 875–885 (1992).
- Montefiori, D. C. *AIDS Res. Rev.* **3**, 161–182 (1993).

A crustal life preserver

Richard W. Carlson

As a raft floating on the underlying convecting mantle, oceanic crust is doomed to sink back into that mantle in less than a couple of hundred million years. Continental crust, on the other hand, survives at the Earth's surface for 4 billion years. Data presented by Reisberg and Lorand on page 159 of this issue¹ provide additional evidence that the mantle directly beneath continental crust exhibits similar survival skills and contributes to the long-term preservation of continental crust on Earth. The results, however, are not only of geological interest, but also demonstrate the potential of a dating technique involving the elements rhenium and osmium.

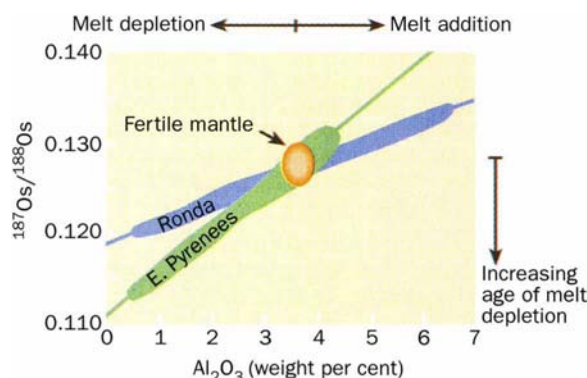
Plate tectonic theory provides a beautifully simple physical explanation for the geological evolution of ocean basins, but is not easily extended to describe the history of the continents (for review see ref. 2). In the oceans, new crust forms by melting of the mantle beneath ocean ridges. As the crust moves away from the ridge, the rigid part of the plate thickens through cooling of the underlying mantle so that it sticks to the overlying plate. After 100 to 200 million years, the oceanic plate has thickened and cooled; its average density exceeds that of the underlying mantle, causing it to founder and sink back into the mantle in the process known as plate subduction.

Ancient continental regions are billions of years old and in some places are underlain by mantle with anomalously fast seismic velocities, indicative of relatively cold temperatures, to depths of several hundred kilometres. Did these mantle 'keels' to the continents form by conductive cooling as in the oceanic case? If so, why haven't they either peeled off from the continent and sunk back into the mantle or caused wholesale continent subduction?

The first clue to this puzzle came from chemical studies that showed the mantle beneath Archaean (>2.6-billion-year-old) continental sections to be melt-depleted compared to both 'fertile' mantle and the mantle beneath ocean basins³. Extraction of partial melt from the mantle removes the aluminium that is required to form the dense mineral garnet, and leaves a mantle residue that is buoyant compared to more fertile mantle. This led to the idea that continents and their adhering mantle are a top chemical boundary layer to mantle convection, rather than the cool,

upper-thermal boundary layer represented by oceanic plates⁴.

The next key question — the age structure of the continental mantle — has proven more difficult to solve, but may now be approachable with the techniques described by Reisberg and Lorand. This question relates directly to the mechanism by which the continental mantle keel forms. Does continental mantle increase in thickness gradually, either by the conductive cooling mechanism involved in ocean basins or by underplating of the



The aluminium versus osmium isotopic correlation determined by Reisberg and Lorand¹ for mantle samples from Ronda, Spain, and the eastern Pyrenees. Fertile mantle, shown by the orange oval, will have Al_2O_3 concentrations of approximately 3.7 weight per cent and an $^{187}\text{Os}/^{188}\text{Os}$ ratio of about 0.127. Removal of melt will decrease the Al concentration from this value and lower the Re/Os ratio of the residue. The longer ago this melt extraction event occurred, the lower will be the present-day $^{187}\text{Os}/^{188}\text{Os}$ ratio of the residue. The linear relationships suggest that the Os isotopic composition of the samples is directly tracking the effects of melt removal and addition shown by the variable Al content. They further indicate that these melt migration events occurred over a relatively short period in each area, but at 1.3 billion years for Ronda and 2.3 billion years for the Pyrenean samples.

residues of continuing magmatic activity on continents? Or are thick sections of melt-depleted mantle formed quickly beneath continental crust by the process that initially formed the crust?

Surprisingly, the difficult part in tackling these issues comes not from the inaccessibility of samples of continental mantle, but in interpreting the isotopic age information in such samples. The problem is twofold.

First, over geological timescales, temperatures in the mantle (generally 1,000 °C or more) allow chemical diffusion continually to homogenize the radiogenic isotopic composition of nearby minerals. So the isotopic compositions of minerals with different ratios of parent to daughter element in a radioactive decay scheme, for example ^{87}Rb which decays to ^{87}Sr , are continually reset. The only effective solution has been to examine minerals

contained within a diffusion barrier, diamond for instance, that isolates the minerals from one another and from their surroundings. Such studies provided the first definitive evidence that some continental mantle is ancient, as old as 3.5 billion years⁵. But diamond inclusions are rare or often absent altogether, and analysing them is tricky and laborious.

The second part of the problem of age determination has to do with the event or events actually being dated by a radiometric system. Until recently, all such systems in use involved so-called incompatible elements, elements that selectively dissolve in the melt when melting occurs in the mantle. On the assumption that continental mantle is the residue of melting, these incompatible elements should be very low in concentration in continental mantle, but they often are not. This is probably because the mantle is continually being subjected to the throughgoing flow of small volumes of melt. Melting of the mantle occurs over a temperature range of several hundred degrees; so if the melting is marginal, only a small volume of melt will be produced but it will be very concentrated in incompatible elements.

These incompatible-element-rich, low-volume melts, if frozen during unsuccessful passage through continental mantle, can leave the continental mantle with its intrinsic deficiency in melt-extractable major elements, such as Al, but cause it to be relatively enriched in incompatible trace elements. Because most of the radiometric systems belong to this group of incompatible trace elements, any age they deliver will be the time of incompatible element enrichment — not the time of the initial melt-extraction event that produced the distinct major element characteristics of continental mantle.

The technique employed by Reisberg and Lorand¹ uses the decay of ^{187}Re to ^{187}Os . Although Re is mildly incompatible, Os is strongly compatible during mantle melting. This means that Os will be low in concentration in melts, but enriched in the residues of melting. So continental mantle should have a low Re to Os ratio that will result in retarded ingrowth of ^{187}Os with time in comparison to fertile mantle. The first application of the Re–Os system to mantle samples from beneath the ancient crust of southern Africa proved this to be the case⁶. In their new study, Reisberg and Lorand demonstrate that the $^{187}\text{Os}/^{188}\text{Os}$ of mantle samples from both Ronda, Spain, and the eastern Pyrenees is well correlated with the Al content of the samples (see figure). This correlation shows that the Re–Os

radiometric system can track the events that determine the major element characteristics of the mantle.

On the assumption that the extrapolation of the observed trends to zero Al content will give the Os isotopic composition of material where Re also has been completely removed, Reisberg and Lorand calculate an age of 1.3 billion years for Ronda and 2.3 billion years for the Pyrenean samples. These ages are in reasonable agreement with the oldest ages obtained for crustal basement rocks from these two areas, thus indicating a temporal and suggesting a mechanistic relationship between crust formation and melt-depletion of the mantle. Similar conclusions have been derived from Re–Os studies of mantle samples from beneath the Archaean crustal sections of Siberia and southern Africa^{6,7}. Archaean regions, however, often are considered anomalous because of the presumed existence of a very high temperature mantle, and hence very high degrees of partial melting, early in Earth history that would produce anomalously buoyant mantle residues.

The results of Reisberg and Lorand show that preservation of melt-depleted mantle roots to the continents is not restricted to the Archaean, but continued throughout most of Earth history. One of the next tasks will be to incorporate the demonstrated existence and longevity of continental mantle keels into geodynamic models for mantle convection, and for the mechanism and formation of continental crust. □

Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington, DC 20015, USA.

1. Reisberg, L. & Lorand, J.-P. *Nature* **376**, 159–162 (1995).
2. Molnar, P. *Nature* **335**, 131–137 (1988).
3. Boyd, F. R. & Mertzman, S. A. in *Magmatic Processes: Physicochemical Principles* (Geochem. Soc. Special Publ. 1 (ed. Mysen, B. O.)) (Geochemical Society of America, University Park, 1987).
4. Jordan, T. H. *Rev. Geophys.* **13**, 1–12 (1975).
5. Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198–202 (1984).
6. Walker, R. J., Carlson, R. W., Shirey, S. B. & Boyd, F. R. *Geochim. cosmochim. Acta* **53**, 1583–1595 (1989).
7. Pearson, D. G. et al. *Geochim. cosmochim. Acta* **59**, 959–977 (1995).

GUTS

How to be physiological

Jared M. Diamond

It is not easy to measure rates of molecular and cell processes in a living animal. So most biochemical and physiological studies are carried out *in vitro*, glossing over the risk that results might differ greatly from conditions *in vivo*. Uhing and Kimura have now succeeded, by ingenuity and technical virtuosity, in devising two new methods for studying small-intestinal transport and metabolism in unanaesthetized unrestrained animals^{1,2}. Their methods have potentially wide applications and the results to date mean that our understanding of intestinal physiology is already having to be rewritten.

Standard *in vitro* preparations for studying intestinal transport include Ussing chambers, everted sacs and sleeves, and membrane vesicles. The accepted 'gold standard' of *in vivo* comparison for these *in vitro* preparations has been a perfused intestinal loop of an anaesthetized animal. Uhing and Kimura's methods instead rest on the principle that arterial blood circulating through the small intestine leaves by the portal vein and passes through the liver before entering the general circulation, the aorta included. Hence concentrations of solutes undergoing absorption are raised in the portal vein compared to the aorta. Kimura *et al.*³ had previously developed techniques for permanently implanting three cannulas simultaneously into a rat's portal vein, aorta and gastroin-

testinal tract, using an introducer needle to place each cannula without interrupting blood flow, and sealing the cannula into place with cyanoacrylate glue.

To translate aortic concentrations of a radiolabelled absorbed solute into an absorption rate, Uhing and Kimura employ a clever dual-infusion method: the solute whose absorption is to be studied is labelled with one isotope (for example ³H) and infused as a bolus into the intestine, while the same solute labelled with a different isotope (for example ¹⁴C) is steadily infused into the portal vein. The ratio, for the two isotopes, of their rates of increase of aortic concentration, multiplied by the known portal vein infusion rate of the ¹⁴C isotope, yields the unknown intestinal absorption rate of the ³H isotope. Alternatively, the ratio of absorption rates of two solutes can be calculated by labelling them with different isotopes, infusing them together into the intestine, and measuring their ratio of portal venous minus aortic concentration difference.

The authors validated their dual-infusion method by showing that their measured rate of solute appearance into the aorta equals the rate of solute disappearance from the lumen of a perfused loop, measured simultaneously in an anaesthetized rat (the gold standard method). But the new technique has four big advantages over the perfused loop

method: the rat is normally unanaesthetized; it is unrestrained and unstressed; its intestine has not been recently manipulated surgically; and it survives for several months, meaning that it can be re-used as its own control for dozens of sequential experiments.

In rats with cannulas already permanently implanted, Uhing and Kimura then went through the traditional anaesthetic and surgical procedures involved in perfused loop experiments. The object was to see how intestinal absorption, as measured by the dual-infusion technique with the transported but non-metabolized sugar 3-O-methylglucose (3-O-MG), was affected. It turned out that anaesthesia alone, without even handling the intestine, depressed absorption by 27%. Anaesthesia plus handling the intestine (as in preparing a loop) decreased absorption far more, by 86%, probably by causing local arterial spasm, decreased portal venous blood flow, intestinal anoxia and decreased intestinal energy supply. Four hours after surgery, absorption was still 62% depressed and did not recover until the next day. So even the best traditional method of measuring intestinal absorption rates *in vivo* greatly underestimated them.

From previous studies in perfused loops with perfused blood supply, it had seemed that most D-glucose absorbed from the intestine is converted to lactate in the course of absorption⁴. However, 98% of D-glucose proves to be absorbed intact in permanently catheterized, unanaesthetized rats⁵. Experimental hypoxia increased glucose conversion to lactate more than tenfold, suggesting that hypoxia may explain why conversion is artefactually so high in perfused loops.

Another artefact uncovered in perfused loop studies was the claim that most D-glucose absorption is passive (occurring, for instance, by solvent drag), rather than active transport mediated by carriers⁶. The passive and active components can be resolved by using the stereoisomer L-glucose as a measure of passive absorption and 3-O-MG as a measure of active transport. The percentage of total absorption that is passive does increase with increasing sugar concentration, as one would expect because the concentration dependence of passive absorption is linear but that of active absorption is saturable. However, even at the unphysiologically high luminal sugar concentration of 400 mM, more than 94% of sugar absorption still proves to be active. The erroneous conclusion about a relatively large passive component arose because anaesthesia and surgery in perfused loops greatly decreased the active component without affecting the passive component.

Finally, still another hypothesis was that active transport of sugar greatly

increases passive sugar absorption by opening tight junctions between cells⁶. In fact, it turns out that active transport has no physiologically significant effect on passive permeability: tracer L-glucose absorption is the same in solutions of the actively transported sugar 3-O-MG as in the non-transported sugar L-glucose, and is unaffected in 3-O-MG solutions by the active transport inhibitor phlorizin. Studies on intestines of unanaesthetized humans who have swallowed a tube permitting intestinal perfusion yield essentially the same conclusion⁷.

These experiments demonstrate how essential it is to make measurements *in vivo* under truly physiological conditions, and are otherwise doubly heartening. First, they make plain that even in classical areas of science there is still scope for ingenuity. Uhing and Kimura's dual-infusion method could have been put into practice many decades ago, but no one

had thought of the idea. Second, although the surgical techniques of Kimura *et al.*³ are undeniably difficult and require practice, they show that the advertising slogan "We try harder" can be as successfully applied in science as in the car-rental business. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024-1751, USA.

1. Uhing, M. R. & Kimura, R. E. *J. clin. Invest.* **95**, 2790–2798 (1995).
2. Uhing, M. R. & Kimura, R. E. *J. clin. Invest.* **95**, 2799–2805 (1995).
3. Kimura, R. E., LaPine, T. R. & Gooch, W. M. *Pediat. Res.* **23**, 235–240 (1988).
4. Nicholls, T. J., Leese, H. S. & Bronk, J. R. *Biochem. J.* **212**, 183–187 (1983).
5. Rich-Denson, C. & Kimura, R. E. *Biochem. J.* **254**, 931–934 (1988).
6. Pappenheimer, J. R. & Reiss, K. Z. *J. Membr. Biol.* **100**, 123–136 (1987).
7. Fine, K. D., Santa, A. C., Porter, J. L. & Fordtran, J. S. *Gastroenterology* **105**, 1117–1125 (1993).

QUASARS

Hosts of possibilities

J. B. Hutchings

ON page 150 of this issue¹ Disney *et al.* describe observations with the Hubble Space Telescope (HST) that reveal details of the host galaxies of four low-redshift quasars. These results are of note for several reasons. First, we have long sought clear evidence as to the nature and environment of quasars, the most energetic objects in the Universe. Second, there have been high hopes that the spatial resolution of the HST would be an important imaging tool in these studies. Finally, the results of the HST observations have been subject to a range of interpretations, some different from those of Disney and colleagues.

It is widely assumed that quasars are accreting massive black holes at the centres of galaxies, although some other ideas continue to be discussed (for instance that they represent compact starbursts or even non-cosmological redshifts). There is a huge base of observation and theory which bears on many forms of galactic nuclear activity, including Seyfert galaxies, radiogalaxies and quasars, and BL Lac objects. There is also persuasive evidence for 'unification' of some of these as axisymmetric systems viewed from different directions.

The spectra of quasars reveal associated stellar-processed material in a wide range of physical states. The continuum of properties from quasars to lower energy Seyfert galaxies and radiogalaxies also strongly suggests that they are galactic nuclei. But direct evidence for the host galaxies of quasars rests on being able to detect and measure properties of such

galaxies. If normal they must be very faint, extending a few arcseconds around the quasar nucleus, which is usually considerably brighter at any observable wavelength. The parallels between radio-loud and radio-quiet galaxies with lower nuclear activity have predisposed us to expect that the hosts of radio-loud and radio-quiet quasars are elliptical and spiral galaxies, respectively.

The principal difficulty is in removing the scattered and diffracted nuclear light from the host-galaxy image. The most important quantity to minimize is the angular size of the point-spread function (PSF; the exact shape of the image of a point source), and clearly this is where the HST scores because it is free from atmospheric turbulence. But there are several other parameters where it does not: signal level, detector noise, and PSF stability and sampling. Although I am a strong supporter of the HST, I am troubled by aspects of the way new HST data are sometimes presented. Many 'first detection' claims are made which in truth are simply improvements on discoveries made elsewhere; and there is an assumption that HST data automatically supersede or out-date earlier investigations. Presentations of the results of quasar imaging with the optically corrected HST have suffered from both these effects, and I would like to note two things: the extent of our ground-based knowledge, and the limitations as well as the excellence of the HST data.

Quasars have been resolved with ground-based telescopes for over 20

years, with detectors that go back to bare photographic emulsions. There have been numerous studies with 4-m-class telescopes, which have steadily improved with low-noise, high-quantum-efficiency detectors, and various levels of active optics. These studies have resulted in resolved images of some 200 quasars, at redshifts ranging from 0.1 to more than 2. Improved capability has continued to reveal more detail about the quasar hosts, and has not led to the emergence of a recognized class with no detectable host galaxy. I consider that there is no reasonable doubt whatever over the existence of host galaxies, and hence of the basic premise, from this large body of work.

The ground-based studies have gone further. Many of the host galaxies are seen to be in tidal interaction with another, usually smaller, companion (ref. 2 and references therein). Luminosity profiles have been derived and many of them are peculiar. Large clouds of emission-line gas have been found in and around quasar host galaxies. Clearly, nuclear activation is often associated with galaxy-wide disturbances. Unexpectedly, quasars at redshift larger than 2 have been resolved, indicating extraordinary ultraviolet rest-frame luminosity and cosmic evolution³. And the best-resolved, lower-redshift objects show signs of compact bright features within the host galaxies. This, then, is the background to the new HST imaging.

The HST images have unprecedented resolution and thus offer an exceptional opportunity to further these studies. But as Disney *et al.* point out, the job is not easy. First, and worst, the instrument with the necessary dynamic range, the Wide-Field Planetary Camera 2, has a spatially and temporally variable PSF, with complex structure. This is caused by the shadowing of the detector planes by optical supports, which is different at every pixel. The telescope undergoes continuous small distortion ('breathing') around its orbit, also altering the PSF. Significant details of the PSF vary over a few pixels, and no model has yet matched the observed structure. So PSF removal cannot be done precisely — in particular, the strong principal diffraction spikes cannot be used, and Disney *et al.* rightly choose to ignore them.

Second, the PSF structure, and indeed the images themselves, are undersampled by the camera: both the 0.1"-sampled wide-field and the 0.046"-sampled planetary camera. There is significant scattered light in both, possibly somewhat worse in the better-sampled planetary camera. And better sampling naturally involves lower signal, which must compete with 'read' noise. Finally, compared with a 4-m-class ground-based one, HST is a small telescope with low throughput: at a spot in one host galaxy well outside the PSF, I have a signal in an image, taken

Broken halo reveals all

W. Gelletly

using the Canada–France Hawaii Telescope, which is four times the read noise. At the same spot in my HST planetary-camera image with the same exposure, the signal is 25 times lower, at a marginal 1/6 of the read noise.

Three approaches to these problems have been used to date. Bahcall *et al.* have used the wide-field camera and subtracted PSFs to minimize the main diffraction spikes^{4,5}. My colleagues and I have attempted deconvolution of the HST images, both on their own and combined with the best ground-based data⁶. Disney *et al.* have ignored the diffraction spikes and discussed azimuthally averaged profiles with model PSFs. In fact we have all done more than this, but it is becoming clear what we can and cannot learn reliably from HST imaging. HST is a poor telescope for detecting faint and outer structure in quasar host galaxies: non-detection is not proof of non-existence. But HST is unsurpassed at detecting details of bright features near to quasars.

The derived luminosity profile of a host galaxy is sensitive to the way the PSF is removed, but is most likely to be reliable in the 1–3'' radius. What we have learned from results of Disney *et al.* and others is that in this range host galaxies are generally structureless at the 0.05'' level, with the exception of bright compact features placed irregularly within them. Further out, the profiles from ground-based data do not all support the simple dichotomy between disk and bulge galaxies; nor indeed do the presence of the bright features. These features range from unresolved knots which could be compact clusters of hot stars, to nuclei of merging galaxies, to details of tidal tails. Other possibilities exist, for instance that the features are dust lanes or optical jets.

What is remarkable is their prevalence: with more observing time, more careful observing strategies, different filters and future instruments, we will use the HST to learn about some of the processes that fuel the nuclei, and how they differ from non-‘active’ interacting galaxies. Observations from HST also offer the opportunity to learn the timescales and lifetimes of the nuclear events, by understanding the ancillary features we can now resolve. □

J. B. Hutchings is at the Dominion Astrophysical Observatory, 5071 West Saanich Road, RR 5 Victoria, British Columbia V8X 4M6, Canada.

1. Disney, M. J. *et al.* *Nature* **376**, 150–153 (1995).

2. Hutchings, J. B. & Neff, S. G. *Astr. J.* **104**, 1–14 (1992).

3. Lehnert, M. *et al.* *Astrophys. J.* **393**, 68–80 (1992).

4. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. *Astrophys. J.* **435**, L11–L14 (1994).

5. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. *Astrophys. J.* (in the press).

6. Hutchings, J. B. & Morris, S. C. *Astr. J.* **109**, 1541–1545 (1995).

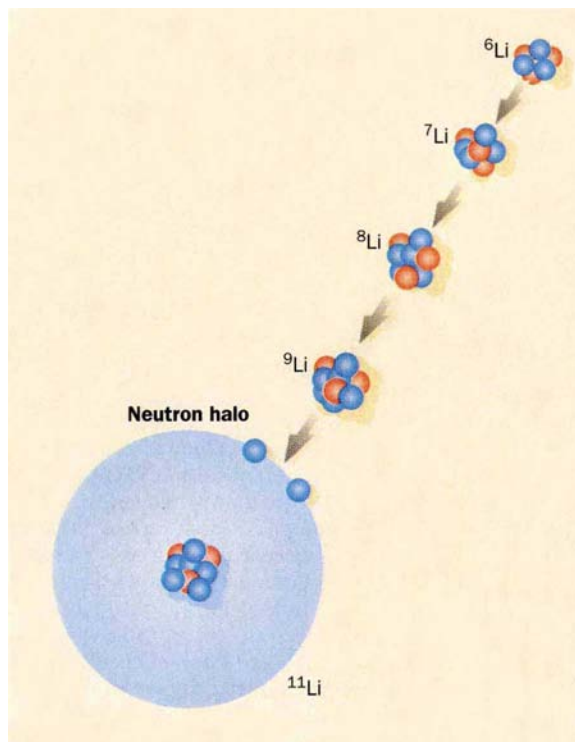
HALOES, it seems, are not just for saints and angels but for some species of atomic nuclei as well. Unlike the celestial haloes of the heavenly host, nuclear haloes form a cloud of neutrons extending well beyond the surface of a tightly bound core of neutrons and protons. Only a few halo nuclei are known, all of them very light and with an excess of neutrons over their stable cousins, which means that they are radioactive. In a paper in *Physical Review Letters*¹, Bazin *et al.* now report that ^{19}C , which has seven more neutrons than the most stable carbon nucleus ^{12}C , is a halo nucleus, the heaviest yet observed.

Why is it that only now are we finding

nuclei at high energy, the projectile nucleus gains or loses a significant number of protons and neutrons. This residual nucleus, unstable and with an excess of protons or neutrons compared with the stable species, moves forward with the momentum of the original beam particles. As a result these nuclei form a beam which can be used to initiate and study nuclear reactions. These radioactive nuclear beams are weak in intensity and poorly defined in energy, but sufficient nuclei are produced to allow their properties to be studied, as in the case of ^{19}C .

There is now a variety of evidence to show that some nuclei have haloes, involving either one or two neutrons. The first indications came from Tanihata *et al.*², who showed that the total interaction probability for such nuclei passing through a thin target was much larger than one would expect from their geometrical size, if it followed the established prescription for stable nuclei. In other words they were much larger than expected. Experimenters then studied the breakup of halo nuclei. For example Bazin *et al.* examined the breakup of ^{19}C into ^{18}C plus a neutron in the nuclear or Coulomb field of another nucleus. They measured the distribution in momentum of the residual ^{18}C core particles. The measured spread in momentum of these particles is directly related to their initial spatial distribution in ^{19}C via Heisenberg's uncertainty principle. This tells us that a narrow spread in momentum reflects a large spread in space.

If one measures the spread for ^{18}C particles at right angles to the beam there is a further complication because the spread is distorted by their subsequent deflection by other nuclei as they leave the target. Bazin *et al.* get around the problem by measuring the ^{18}C momentum in the forward direction where there is no deflection. By definition, if the particles go straight forward they have not been deflected after they were produced. From these measurements one can deduce the initial distribution of neutrons and the result is that the ‘halo of ^{19}C extends far beyond the ^{18}C core’¹. This result is confirmed by the



Nuclei of the element Li, all of which have three protons plus an increasing number of neutrons. As neutrons are added, the neutrons and protons occupy the same volume until the two-neutron halo nucleus ^{11}Li is reached; here the last two neutrons form a halo and there is a sudden jump in the radius of the nucleus.

such nuclei? The reason is that almost all of our knowledge of the properties of atomic nuclei has come from studies of the radiation and particles emitted in collisions between nuclei. Because nuclei have positive electrical charges it is necessary to accelerate at least one of them to high velocity to overcome the electrical repulsion between them. Until very recently it has only been possible to accelerate the 283 stable or very long-lived nuclear species found on Earth.

This barrier is now crumbling. In some peripheral collisions between two heavy

results of a measurement of the distribution of neutrons emitted following ^{19}C breakup. The latter experiment, carried out by Angelique *et al.* at GANIL, the heavy ion accelerator at Caen in France, was reported at a meeting in June³.

A variety of nuclear models has been used to explain the properties of halo nuclei. The essence of these models is that we have a quantum mechanical system, one in which neutrons and protons are bound together in a potential well. In stable nuclei the last proton or neutron is bound by about 8 MeV, quite deep in the well. In halo nuclei the binding of the last neutron or neutrons is only a few per cent of this. In the everyday world this would still mean that the neutron would stay within range of the force that binds it to the nucleus. Because it is loosely bound it will not take much energy to move the last neutron in ^{19}C a long way from the ^{18}C core. In a quantum system, the uncertainty principle means that it can stay there for a comparatively long time. Hence the outermost neutron in ^{19}C spreads out in a cloud around the core.

The properties of halo nuclei depend critically on the pairing attraction of the most loosely bound neutrons. This will dictate where the nucleus can no longer hold another neutron; it leaks out of the potential well immediately. This defines the neutron 'drip line', which is quite unknown. In heavier nuclei the halos are expected to give way to a neutron skin, a thin envelope of neutrons outside the core nucleus. Although these skins are less dramatic than haloes, they will cause many changes in nuclear properties. One may expect the shell structure observed in stable nuclei to change drastically or disappear, new types of excitation at low energy due to the relative motion of neutrons and protons, and a different behaviour in nuclear reactions.

These results on neutron haloes are just the first fruits of experiments with radioactive nuclear beams. Beams from fragmentation processes will remain limited in intensity but an alternative, the two-step method, promises intense beams. One accelerator is used to create the radioactive nuclei in a target; the nuclei are then stopped, ionized and injected into a second accelerator. A simple system of this kind is in operation at Louvain-La-Neuve based on a small cyclotron. It works well but can produce only a few nuclear species. It has already been used to study proton capture on ^{15}N and ^{19}Ne , reactions of considerable astrophysical importance. Two larger and roughly comparable facilities are now under construction in Europe, at GANIL and CERN, which will provide a wider range of useful beams.

Slightly further in the future, a full-scale, second-generation facility could be built at the Rutherford Appleton Labora-

tory in Britain, which would use 100 μA of 800 MeV protons from the ISIS synchrotron to create the radioactive nuclei in a heavy target. Following ionization they would be accelerated by a new linear accelerator. Tests of whether a suitable ion source can be built are now in preparation⁴. Later this year it is hoped to begin a full design study and costing for this facility.

These new beams lie in the future. They will create opportunities in nuclear astrophysics, condensed matter studies and medical applications, and open up the whole range of atomic nuclei to study. Meanwhile experiments such as that by

SOLAR PHYSICS

Waves in the wind

Douglas Gough

THE spectrum of the temporal variation of particle fluxes in the interplanetary plasma is dominated by sharp lines. This truly remarkable phenomenon is reported by Thomson, MacLennan and Lanzerotti on page 139 of this issue¹. The discovery is amazing in the light of current thinking about the solar wind, because such thinking is based on the idea that the temporal variation is predominantly the consequence of turbulence, which has a relatively smooth spectrum.

So where do the periodicities come from? The hypothesis put forward by Thomson *et al.* is that they are a manifestation of the coherent seismic oscillations of the Sun. If true, that would surprise both helioseismologists and space plasma physicists, because more or less everyone has always assumed that any seismic disturbance propagating through the chromosphere into the corona would be washed out by the substantial inhomogeneities in the upper solar atmosphere. It would also delight the helioseismologists, for it would represent the discovery of solar g — or gravity — modes. If these modes could be identified, they would put very helpful constraints on the possible structure of the Sun's interior, which in turn would bear on our understanding of stellar evolution and the solar neutrino problem.

The new results come from a spectral analysis of fluxes of H and He ions measured on the Ulysses and Voyager 2 spacecraft. One-hour averages were available from nine separate channels of data — eight from Ulysses spanning intervals of 70–245 days in 1992–94, an epoch of moderate solar activity, and one from Voyager spanning more than 300 days mostly in 1985. To avoid confusion from bias caused by gaps in the data, a quite sophisticated analysis of the data was carried out. The upshot was a strong

Bazin *et al.*¹ not only tell us that neutron haloes persist into heavier nuclei but hint at the full promise of science with radioactive nuclear beams. □

W. Gelletly is in the Department of Physics, University of Surrey, Guildford, Surrey GU2 5XH, UK.

1. Bazin, D. *et al.* *Phys. Rev. Lett.* **74**, 3569–3572 (1995).
2. Tanihata, I. *et al.* *Phys. Lett.* **B160**, 380–384 (1985).
3. Angelique, J. C. *et al.* *Proc. Int. Conf. on Exotic Nuclei and Atomic Masses* (ed. Guillemaud-Mueller, D.) (Editions Frontières, Arles, France, in the press).
4. Bennett, J. R. J. *et al.* *Proc. Workshop on Production of Intense Radioactive Beams at the Isospin Lab.*, 289–292 (CONF-9210121, Oak Ridge Natn. Lab., Tennessee, 1992).

statistical case against misinterpretation of the results. The frequencies of the spectral lines from the different channels were then compared, and were assigned to a group if they agreed within 0.2 μHz , each member of a given group being regarded as a separate measurement of the same oscillator. There are 35 groups containing measurements from at least four channels, the mean frequencies of which are tabulated in the paper by Thomson *et al.*¹, and a further 126 groups with three members. The frequencies range from 0.446 μHz , corresponding to the characteristic period of rotation of the Sun, to the Nyquist frequency 139 μHz of the 1-hour sampling.

It is hard to imagine how coherent oscillations can be intrinsic to the inhomogeneous, temporally fluctuating solar wind. Therefore it is natural to conjecture that the source is the Sun. Aside from rotation, the only strictly periodic phenomena are the seismic oscillations, the only candidates for most of the frequencies found by Thomson *et al.* being g modes. Although hints of such modes have been reported in the past from optical helioseismic data, no really convincing detection has yet been made. It now appears that, serendipitously, the Ulysses and Voyager particle collectors are the most sensitive seismometers for the task.

The credibility of the hypothesis that helioseismic modes have been detected would be greatly strengthened if modes that have previously been unambiguously identified in the optical observations could be seen also in particle-flux data. With this in view, Thomson *et al.* investi-

1. Thomson, D. J., MacLennan, C. G. & Lanzerotti, L. J. *Nature* **376**, 139–144 (1995).
2. Elsworth, Y. *et al.* *Astrophys. J.* **434**, 801–806 (1994).
3. Libbrecht, K. G., Woodard, M. F. & Kaufman, J. M. *Astrophys. J. Suppl. Ser.* **74**, 1129–1149 (1990).

gated energetic electron-flux data from Ulysses to try to find frequencies of solar p (pressure) modes — these are acoustic standing waves. The relevant data have a 1-minute resolution, and are adequate to hold the signature of 5-minute acoustic oscillations. It should be appreciated that, unlike the ion-flux analysis, these data were studied with the objective of testing a specific hypothesis, so a positive outcome would be strong confirmation.

That outcome was realized. Of the 118 modes with degrees up to six in the frequency range 1,588–4,200 μHz observed by the Birmingham Solar Oscillations Network (BISON)² and the Big Bear Solar Observatory³, 90 were matched to within 1.2 μHz by the electron-flux data; the latter were systematically higher by a factor 1.00078. When the Ulysses data were scaled down by that factor, the mean discrepancy with the BISON data² at low solar activity for modes of orders 15–25, whose frequencies are amongst those most precisely determined, is only 0.68 μHz . This is comparable with the apparent frequency modulation arising from beating between unresolved rotationally split modes.

Granted that the interpretation is correct, several issues remain. The most obvious is that of how solar oscillations of tiny amplitude can induce measurable particle-flux variations. Theoretically, the amplitude of the relative-density fluctuation increases with height through the solar atmosphere, and could propagate with yet increasing amplitude through the solar wind. It would certainly be interesting to calculate in detail the propagation of the disturbance in order to compare the particle-flux amplitudes with the optical data. In particular, the g -mode amplitudes might be great enough to be detected gravitationally by instruments such as the Laser Interferometer Space Antenna, a suite of spacecraft currently under study by the European Space Agency. Moreover, having a substantial lateral component of motion, g modes could modulate the interplanetary magnetic field, which Thomson *et al.* have confirmed from spacecraft data. The factor of 1.00078 between the particle-flux and optical frequencies was described by Thomson *et al.* as possibly being a consequence of the Doppler effect. An alternative possibility is that it is a product of variation associated with the solar cycle: the Ulysses data were obtained during moderate solar activity, when frequencies should be about 0.2 μHz greater than those near the solar minimum² with which they were compared. It would be worth trying to find a temporal trend in the Ulysses data.

Perhaps of greatest interest is the diagnostic role that g modes play in helioseismology. These modes predominantly reside deep in the radiative interior, and carry a great deal of information about the

energy-generating core of the Sun. To be of use, any mode whose frequency has been measured must be identified. Thomson *et al.* have taken a step towards that goal by looking for phase jumps across eigenfunction nodal surfaces as the Ulysses spacecraft traversed heliographic latitude. Perhaps complete success will come when the data are coupled with imminent observations from the ground-based Global Oscillations Network Group of helioseismic observatories, and from the

seismic instruments soon to be launched on the SOHO spacecraft. Then we might learn whether conditions at the site of the production of solar neutrinos are as theories of stellar evolution would have us believe. $\square$

Douglas Gough is in the Department of Applied Mathematics and Theoretical Physics, and the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

EVOLUTIONARY THEORY

Mutualism on the move

Peter Hammerstein and Rolf F. Hoekstra

THE modern picture of evolution is one in which natural selection tends to operate more effectively at the level of individuals than at the level of groups, species or higher units. Inter-individual competition thus seems to be the key to understanding most of the organismic design. In this view of the living world, conflict seems very natural and cooperation appears as a phenomenon that usually requires subtle explanation. Such a need for sophistication was recognized early on by Hamilton and Trivers^{1,2}, who based their theories of cooperation on genetic relatedness and on the logic of repeated interactions. Much of the subsequent theoretical work on the evolution of cooperation is closely related to these two approaches — so closely in fact that the field began to stagnate. At an IIASA workshop* on mutualism and cooperation, a new 'migratory restlessness' in the minds of theoreticians was evident.

This innovative tendency relates strongly to a growing interest in two areas of research: cooperation between individuals of different species and cooperation between different components of an organism's genetic material. In the former case, the interacting units do not share a common gene pool; in the latter, this may or may not be the case. When there are separate pools, the interacting parts do not share genes that are identical by descent. The modeller is then forced to leave the classical paradigm of kin selection theory.

A weak flavour of this theory is retained, however, if one looks at various possibilities for nonrandom associations of 'complementary genes' in interacting partners. Here a specific allele of gene A in one species is particularly likely to meet a specific allele of gene B in the other. What can cause such an association? This question leads straight into the analysis of

spatial population structure (E. Szathmáry, Eötvös Univ., Hungary).

We also think that the study of long-term cooperation is undergoing fundamental change. The notion of a repeated game has been the modeller's traditional frame of reference, with attention heavily focused on the Prisoner's Dilemma. A repeated game is a sequence of similar interactions between the same partners which ends as a result of an external event. For models of this kind, classical game theory provides the so-called folk theorem which roughly states that any degree of possible cooperation can be found in an appropriate Nash equilibrium if the game lasts long enough. Obviously, the repeated Prisoner's Dilemma looks very special when compared with the immensely large universe of models to which the folk theorem applies. This particular game seems to have created some paradigm blindness so that the mainstream literature realized neither the full scope of the folk theorem nor a very important limitation of the underlying model structure.

To see the limitation of repeated games, look at the following example. In the mutualistic relationship between a cleaner fish and its regular customer, the customer might defect from cooperation by swallowing the cleaner while it is acting as a living toothbrush. This would clearly end the game because a dead cleaner cannot act any more. Instead of an external event, the defection from cooperation by one of the players causes the interaction sequence to end. Therefore, we are not facing a repeated game. This is not a minor technical point, because the basic character of the game is changed and a different analysis of the problem is required. Instead of restricting attention mainly to the binary interaction, it becomes essential to ask what gains could be achieved by trade with other partners. In the cleaner example, this means a study of the local cleaner market.

Returning to the spatial structure, it has

*The Evolution and Ecology of Mutualism, International Institute for Applied Systems Analysis, Laxenburg, Austria, 12–15 May 1995.

long been recognized that in a 'viscous' population with interaction between neighbours, kin selection will operate. Something similar can also occur in interactions between species. But more subtle and unanticipated effects have been found using cellular automata in theoretical explanations of the coexistence of autonomously reproducing entities. P. Hogeweg (Univ. of Utrecht), who is a pioneer in using this approach to study the stability of prebiotic chemical cycles, reported on the emergence of important mesoscale spatial patterns. These patterns may stabilize hypercycles of cooperating molecules. T. Czárán (Eötvös Univ.) simulated a system of different types of self-replicating molecules in which reproduction requires the presence of all types. Owing to frequency-dependent advantage to rare types, this system can be stable when interactions are within a limited neighbourhood, whereas stability is lost in a spatially homogeneous population.

Non-cooperative 'cheating' variants are considered to be the major obstacle to the evolutionary maintenance of cooperation. As emphasized by J. Bronstein (Univ. of

Arizona), the empirical evidence shows the presence of cheaters in almost every mutualistic system. This means on the one hand that theoreticians correctly emphasize the possibility of cheating. On the other, it also demonstrates that cheating does not necessarily destabilize the mutualism. It must also be said that the study of mutualism involves more than stability analysis. One area where theoretical research is still very limited is the evolution of various forms of mutualism, starting from non-mutualistic interactions. Although the classical view that parasitism should evolve to commensalism and ultimately to mutualism has been recognized as too simplistic, it is important to understand when such evolution would be expected to occur. N. Yamamura (Saga Medical School, Japan) has analysed a host-parasite model in which the parasite first wins the conflict with the host about vertical transmission, and then evolves reduced virulence. A similar explanation may hold for the evolutionary origin of cell organelles.

Finally, it was clear from the meeting that cooperation can occur in rather dif-

ferent situations. Common interest is an obvious incentive for cooperation but is far from being sufficient. For example, it is in the common interest of two parents that their offspring should be raised successfully. Who would argue, however, that this implies a shared interest in paying the price of parental care? Cooperation resulting from common interest is, therefore, expected not to be very common and it is more likely within a species than between species. But now consider a situation in which two individuals are trading different commodities that correspond to different needs. The basic asymmetry here suggests that cooperation based on trading may evolve more often between species than within species. □

Peter Hammerstein is at the Max Planck Institute for Animal Behaviour Research, D-82319 Seewiesen, Germany. Rolf F. Hoekstra is in the Department of Genetics, Wageningen Agricultural University, NL-6703 HA Wageningen, The Netherlands.

1. Hamilton, W. D. J. *theor. Biol.* **7**, 1–16, 17–52 (1964).
2. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).

EPILEPSY

Taking up GABA again

Brian Meldrum

THE neurotransmitter GABA (γ -aminobutyric acid) was identified some 35 years ago as the principal inhibitory transmitter in the brain. As compounds that block the synthesis, synaptic release or postsynaptic action of GABA also induce convulsions, many researchers have asked whether features of the natural types of epilepsy might be explained by some alteration of the inhibitory effects of GABA. Because GABA-containing neurons and GABA receptors are preserved in temporal-lobe epilepsy¹, During *et al.* (page 174 of this issue²) explore the possibility that it is a shortfall in the amount of GABA released that is to blame for the loss of inhibitory function. They find that the number of GABA transporters is reduced in epileptic hippocampus, and suggest that release of GABA during seizures is by reversal of GABA transport, and therefore that the amounts of GABA released are insufficient to suppress seizure activity.

GABA release from neurons is of two types^{3,4}. Classical synaptic release is vesicular, calcium-dependent and sensitive to tetanus toxin, and can be triggered by high K^+ *in vitro* and *in vivo*. There is also non-vesicular, calcium-independent GABA release that is secondary to depolarization of the postsynaptic membrane and Na^+ influx. This release depends on reversal of the GABA transporter and is blocked by GABA-uptake

inhibitors such as nipecotic acid. Glutamate, veratridine and anoxic depolarization cause release of GABA by this mechanism.

During and Spencer had earlier used microdialysis probes implanted in patients to show that there is a greater rise in extracellular GABA concentration in the non-epileptogenic hippocampus than in the epileptogenic hippocampus during seizure⁵. In their latest work² they use high K^+ and glutamate to provoke increases in extracellular GABA concentration and find that vesicular release of GABA is increased, but non-vesicular release is decreased, in the epileptogenic hippocampus compared with the contralateral non-epileptogenic hippocampus. The exogenous GABA extracted from the dialysis probe was relatively reduced on the epileptogenic side. In rats in which limbic seizures were triggered following electrical kindling, the number of GABA transporters was reduced by about 50 per cent. Lack of control material prevents a comparable analysis of GABA transporter density in human temporal lobectomy specimens.

What is the functional effect of a reduced GABA transporter activity? Answers to this are supplied by electrophysiological studies in hippocampal slices employing GABA-transport inhibitors such as nipecotic acid, SKF 89976A

and tiagabine^{6,7}. At A-type GABA receptors, the amplitude or initial decay of inhibitory postsynaptic potentials is little altered but their late decay is augmented; whereas slow inhibitory postsynaptic potentials, due to activation of postsynaptic B-type GABA receptors, are enhanced. This enhancement of GABA-mediated inhibition would be expected to diminish the spread of focal seizure activity. Inhibition of GABA uptake also potentiates more distant actions of synaptically released GABA. Excitatory transmission in the CA1 region in a hippocampal slice can be depressed by GABA acting on presynaptic GABA B-type receptors following tetanic stimulation of inhibitory inputs; this phenomenon is greatly enhanced in the presence of a GABA-uptake blocker.

That the enhanced inhibition associated with blockade of GABA uptake (or with a reduction in the density of GABA transporters) has an anticonvulsant effect is supported by studies in animal models of epilepsy with uptake inhibitors such as tiagabine. Tiagabine is now in clinical trial for treatment of complex partial seizures. A reduced GABA transporter activity can thus be seen as a compensatory or corrective change in and around the epileptogenic focus.

During and colleagues² propose another way in which this change may be significant for the evolution of seizure discharges. They suggest that the release of GABA during seizures may be non-vesicular and that this release may be important for suppressing seizure activity, but this is less certain. The inward trans-

port of GABA is linked to the co-transport of Na^+ and Cl^- (down their concentration gradients); it is electrogenic and therefore influenced by the membrane potential. The depolarization and the inward flux of Na^+ following ischaemia is sufficient to reverse the carrier action of the GABA transporter (and the glutamate transporter). Neuronal depolarization during seizure commonly reaches -40 or -30 mV, but -20 mV is required to reverse GABA transport. There is no direct evidence that the increase in extracellular GABA concentration during seizures⁵ is non-vesicular — the greater increase on the non-epileptogenic side may merely reflect increased synaptic release. The functional significance of increases in extracellular GABA secondary to depolarization and Na^+ gradient failure is also uncertain.

Future developments are likely to exploit recent advances in the molecular biology of GABA transporters. Four GABA transporters (GAT1–4) have been cloned and expressed⁸. GAT1 (also known as GATA) and GAT4 are neuronal, GAT2 is glial. Tiagabine and SKF 89976A are selective inhibitors of GAT1 (ref. 9). It is likely that the changes observed by During *et al.* in the epileptogenic hippocampus relate to changes in GAT1, if only because it is the predominant neuronal GABA transporter in the hippocampus.

The conduct of human biochemical pharmacological studies of a kind more normally performed in animal models has provided some provocative new interpretations of epileptic pathophysiology. Further investigation is urgent, but the window of opportunity may be closing owing to other technological advances. Dramatic improvements in neuroimaging, particularly higher-resolution magnetic resonance imaging scans and the application of magnetic resonance spectroscopy, mean that fewer patients require chronic depth-electrode implantation. Studies of the kind so effectively pioneered by During and Spencer⁵ may no longer be viable. □

Brian Meldrum is in the Department of Neurology, Institute of Psychiatry, and the Institute of Epileptology, De Crespigny Park, London SE5 8AF, UK.

Demise of the Atelocerata?

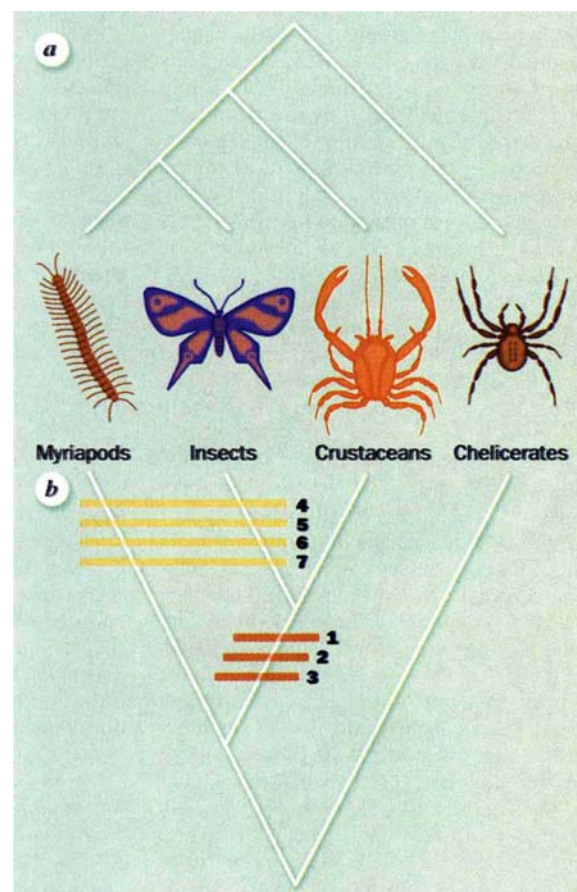
Maximilian J. Telford and Richard H. Thomas

THE relationships among the major lineages of that spectacularly successful group of 'jointed-legged' animals, the arthropods, have provided morphologists, embryologists and palaeontologists with over 150 years of unresolved controversy. The arthropods comprise the insects, crustaceans (crabs, shrimps and the like), myriapods (centipedes, millipedes and allies) and chelicerates (the horseshoe crabs and arachnids) and are probably close to more enigmatic groups such as the lobopods (velvet worms). The difficulties found in discerning arthropod relationships typify the more general challenge to biologists studying relationships between groups arising early in the history of animals. Principal among these difficulties is that natural selection, acting over hundreds of millions of years, sometimes causes related animal groups to diverge morphologically while unrelated ones may converge to become superficially similar. Two papers appearing on pages 163 and 165 of this issue^{1,2}, using disparate sources of molecular data, claim to have some of the answers to the question of arthropod relations and, gratifyingly, they broadly agree.

Almost every possible phylogeny of the arthropods has had its adherents, but only three hypotheses have received wide support in recent decades. The polyphyletic view, championed by Sidnie Manton³, derives three distinct groups of living arthropods from different non-arthropod ancestors. These three groups are the crustaceans, the chelicerates and the Uniramia (the lobopods and the myriapods plus insects, all of which have single-branched legs). The polyphyletic view has recently been widely rejected using both morphological evidence and molecular data. The two other commonly accepted views derive all arthropods from a single arthropod-like ancestor. These are the 'TCC' theory^{4,5}, which has a group

containing the extinct trilobites (T), the crustaceans (C) and the chelicerates (C) separate from the insects and myriapods, and the 'mandibulate' hypothesis (a in the figure), linking the crustaceans to the myriapods plus insects (the Mandibulata)⁶, leaving the trilobites and chelicerates as outgroups.

Environment-driven convergence, a problem inherent in morphology-based phylogenetic analyses of the arthropods, has been considered. Gene sequence data have been widely seen as a panacea for this problem as changes in gene sequences



a, A phylogenetic tree showing the popular mandibulate hypothesis of arthropod relations which links the myriapods and insects with the crustaceans as closest sister-group⁶. b, A phylogenetic tree showing the close relationships between the insects and crustaceans supported by two papers in this issue^{1,2}. Brown bars indicate shared derived morphological traits of the insects and crustaceans which support this link: (1) mandible formed of a fused limb base^{6,10}; (2) detailed similarities in the ontogeny of the nervous system¹²; (3) detailed similarities in the cellular composition of the eyes and their connections to the brain¹³. Yellow bars indicate characters inferred to be shared by insects and myriapods through convergence due to terrestrial life²: (4) uniramous limbs; (5) tracheal breathing; (6) malpighian tubules for nitrogen excretion; (7) lack of appendages corresponding to the second antennae of crustaceans. According to this hypothesis of arthropod relations, both the myriapods and crustaceans might be paraphyletic.

1. Babb, T. L., Pretorius, J. K., Kupfer, W. R. & Crandall, P. H. *J. Neurosci.* **9**, 2562–2574 (1989).
2. During, M. J., Ryder, K. M. & Spencer, D. D. *Nature* **376**, 174–177 (1995).
3. Pin, J.-P. & Bockaert, J. *J. Neurosci.* **9**, 648–656 (1989).
4. Belhage, B., Hansen, G. H. & Schousboe, A. *Neuroscience* **54**, 1019–1034 (1993).
5. During, M. J. & Spencer, D. D. *Lancet* **341**, 1607–1610 (1993).
6. Thompson, S. M. & Gähwiler, B. H. *J. Neurophysiol.* **67**, 1698–1701 (1992).
7. Isaacson, J. S., Solis, J. M. & Nicoll, R. A. *Neuron* **10**, 165–175 (1993).
8. Amara, S. G. & Kuhar, M. J. *A. Rev. Neurosci.* **16**, 73–93 (1993).
9. Borden, L. A. *et al. Eur. J. Pharmac.* **269**, 219–224 (1994).

are thought to be largely unaffected by the environment, with the result that similarity of DNA sequence more accurately reflects relationships. The reality of sequence evolution is more complex than this statement suggests, making difficult the retrieval of phylogenetic signals from such ancient divergences. Friedrich and Tautz² have analysed the largest arthropod DNA sequence data set so far from the small and large subunit ribosomal RNA genes of a carefully selected range of taxa. Importantly, they have also taken pains to test for and control several factors known to bias the recovery of the correct phylogeny.

Boore *et al.*¹ use molecular data in a manner more familiar to traditional systematists. They have looked for gene rearrangements within the small mitochondrial genomes of animals (37 genes, usually) and used gene boundaries as characters to produce a phylogenetic tree. These rearrangements occur only rarely and are considered to be immune to selective pressures. The vast number of potential rearrangements makes identical novel arrangements in unrelated taxa extremely unlikely. Events such as these provide ideal characters for phylogenetic analysis. Both studies provide further strong support for a monophyletic Arthropoda, a concept particularly strengthened by the novelty of the data of Boore *et al.*, who also tentatively include the lobopods in this assemblage.

The one grouping common to virtually all previous schemes relating the arthropods is the close relationship of the myriapods and insects; a group united by the common possession of several adult characteristics and often referred to as the atelocerates. The existence of the Atelocerata is disputed by the study by Friedrich and Tautz, who suggest that it is the crustaceans and not the myriapods that are the sister group of the insects (incidentally placing the origin of the myriapods earlier than their fossil record suggests) (*b* in the figure). Boore *et al.* also present data that support this conclusion. Their interpretation of their data is cautious, however, and they do not separate the traditional atelocerate clades on their summary diagram. The closeness of the insects and crustaceans had been supported by some, but not all, previous molecular analyses⁷⁻⁹.

If the close relationship of the crustaceans and insects is accepted, features shared by myriapods and insects, but not the crustaceans, must be shown either to be shared as a result of convergence, most likely because both are terrestrial groups, or to be primitive characteristics secondarily lost in the Crustacea.

The following characters have traditionally been used to unite the myriapods and insects: unbranched legs, a tracheal system, malpighian tubules, absence of

appendages corresponding to the second antennae of crustaceans, and a mandible composed of a whole limb (crustacean mandibles are formed from a limb base). Friedrich and Tautz argue that the first four of these characters are shared through convergence due to both groups being terrestrial; structures similar to the first three are even found in some terrestrial arachnids. Finally, recent gene expression data argues against a whole limb mandible in the insects. The gene *Distal-less*, generally expressed in the tip of insect appendages, is not expressed in the insect mandible¹⁰. This suggests that the appendage tip is missing and that the insect mandible is, in fact, similar to that of the crustaceans. Indeed, Kukalová-Peck⁶ claims that a limb-base mandible is present in all arthropods.

Other lines of evidence support a sister group relationship between insects and crustaceans¹¹. Nervous system development in the two groups is strikingly similar¹², with no equivalent similarity yet seen in myriapods or chelicerates and, again on a cellular level, crustacean and insect eyes are more similar to each other than either are to those of myriapods or chelicerates¹³.

Despite their convincing congruence, these results are unlikely to be immediately generally accepted and many questions remain. More needs to be done to pin down the position of the lobopods, as well as the positions of other less well known groups such as the tardigrades and pycnogonids (sea spiders). The monophyly or otherwise of both the myriapods and the crustaceans also remains uncertain. The position of the chelicerates is still not clear and will determine whether a revised concept of the Mandibulata survives. It seems unlikely that we will have to wait another 150 years for answers to these questions and with them will come a much improved understanding of the evolution of arthropod body form. □

Maximilian J. Telford and Richard H. Thomas are in the Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

1. Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L. & Brown, W. M. *Nature* **376**, 163–165 (1995).
2. Friedrich, M. & Tautz, D. *Nature* **376**, 165–167 (1995).
3. Manton, S. M. *The Arthropoda: Habits, Functional Morphology and Evolution* (Clarendon, Oxford, 1977).
4. Cline, J. L. *Science* **186**, 13–18 (1974).
5. Willis, M. A., Briggs, D. E. G., Fortey, R. A. & Wilkinson, M. *Verh. dt. zool. Ges.* (in the press).
6. Kukalová-Peck, J. *Can. J. Zool.* **70**, 236–255 (1992).
7. Turbeville, J. M., Pfeifer, D. M., Field, K. G. & Raff, R. A. *Molec. Biol. Evol.* **8**, 669–686 (1991).
8. Ballard, J. W. O. *et al.* *Science* **258**, 1345–1348 (1992).
9. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. *Cladistics* **9**, 1–39 (1993).
10. Panganiban, G., Nagy, L. & Carroll, S. B. *Curr. Biol.* **4**, 671–675 (1994).
11. Averof, M. & Akam, M. *Phil. Trans. R. Soc. Lond.* **B347**, 293–303 (1995).
12. Whittington, P. M., Leach, D. & Sandeman, R. *Development* **118**, 449–461 (1993).
13. Paulus, H. F. in *Arthropod Phylogeny* (ed. Gupta, A. P.) 299–383 (Van Nostrand Reinhold, New York, 1979).

The sap also rises

ALL plants are pumps. They take in water through their roots and transpire it as vapour from their leaves. How do they lift the water? Capillary rise is limited to a metre or so, and an atmospheric suction pump cannot lift water above 10 metres; yet trees can grow 100 metres high. Furthermore, aquatic plants cannot use either mechanism, but still thrive; and even land plants survive flooding.

Daedalus reckons that plants move their internal fluids just as we do: by mechanical pumping. Their internal channels form a distributed peristaltic pump, driven by the ceaseless shaking of the wind. On this view, the gallant waving of a wind-blown field of corn, the ceaseless rustling of forest leaves, even Wordsworth's daffodils "tossing their heads in sprightly dance" have stern biological purpose in their movements.

This theory is supported by a recent finding that many commercial plants, such as tomatoes, aubergines and cucumbers, benefit from the repeated bending of their stems by stroking. They develop a darker green, and grow more compactly, as if the added massage saved them from having to grow extended fronds to catch the wind. DREADCO gardeners are now testing this notion. They are shaking selected plants in a range of frequencies and vibrational modes to discover the most effective pumping regimes. Control plants are being clamped rigidly immobile to see if their growth is stunted. Once the technique has been optimized, vibro-horticulture should speed the sap through crops of all kinds, boosting their metabolism and growth.

For the smaller plants, a standard laboratory shaker should be ideal. Trays of seedlings could easily be vibrated at whatever frequency and amplitude optimized their growth. Very large single plants, such as trees, might also be driven mechanically by ropes or hydraulic rams under programmed control. But for crops such as corn, maize and oats, wind seems the best shaker. A field could be set with big baffles to funnel and direct it, spring-loaded vanes to release a stream of vortices downwind, or huge tuned pipes to resonate at the best frequencies. As the crop matured, these could be adjusted to maintain the optimum pattern of vibration.

Aquatic plants would be easier to vibrate. Daedalus is designing special stirrers for watercress beds, and a central wave machine to spread waves out into a rice paddy-field. He also advocates the planting of seaweed on breakwaters and sea defences. The rougher the waves, the more the weed would grow to damp them. David Jones

Causes of HIV diversity

SIR — The human immunodeficiency virus (HIV) is characterized by enormous genetic flexibility, which gives rise to drug resistance, escape from immune responses and failure of vaccination attempts. There is much discussion about the factors contributing to viral diversity in individual infections. It is clear that the high error rate of the reverse transcriptase¹ and the high turnover rate *in vivo*^{2,3} generate vast numbers of different virus mutants. The diversity of viral quasispecies, however, is shaped by a combination of mutation and selection forces. The main selective forces that have been proposed to drive HIV diversity are the immune response, cell tropism and random activation of infected cells.

To quantify selection pressures on the HIV quasispecies, we have analysed the synonymous (amino-acid preserving) and nonsynonymous (amino-acid changing) nucleotide substitution pattern in the HIV-1 envelope gene of an infected haemophilic patient followed since seroconversion for 7 years^{4,5}. We compared 67 non-identical plasma-derived RNA sequences over a stretch of 231 bases, including the V3 loop and

flanking regions. We derived the mean number of nucleotide substitutions per synonymous site, d_s , and per non-synonymous site, d_n , for all pairwise comparisons of sequences in the samples taken at years 3, 4, 5, 6 and 7 after infection. At year 0 the sequences were completely homogeneous.

The figure shows d_s , d_n , d_s/d_n and the CD4 cell count for years 3, 4, 5, 6 and 7 after seroconversion. As expected, d_s increases with time, due to accumulation of synonymous substitutions. The low value of d_s at year 3 may indicate that a selective sweep occurred in the viral population shortly before this time. It is interesting that d_n initially exceeds d_s , but thereafter increases much more slowly, if at all. At year 3 the d_s/d_n ratio is about 0.1, indicating strong positive selection for amino-acid change. As the infection progresses this selection pressure declines remarkably. By year 7 the ratio is about 1.25, indicating weak negative selection against amino-acid change. The pattern of increase in d_s/d_n agrees well with the pattern of decrease in the CD4 cell count. Whenever the CD4 cell count decreases, the d_s/d_n ratio increases.

A varying selection pressure is compatible with the notion that the immune responses select for viral diversity^{6,7}. Early in the infection the immune system will respond strongly against common viral variants and hence favour rare mutants, thereby providing a strong positive selection pressure for diversification. As the immune system declines, this selection pressure should become weaker.

The pattern of decreasing positive selection, however, is also consistent with the notion that most viral diversity in the V3 region is caused by adaptation for various cell tropisms⁸. At seroconversion the patient carries a strongly homogeneous virus population (in the V3 region), then diversification occurs as HIV infects many different cell types and tissues in the body. Initially this would provide strong positive selection pressure which would decline when the virus has generated many variants with specific cell tropisms.

The observed pattern rejects the hypotheses that HIV diversity in V3 is purely neutral or simply generated by random activation of infected T-cell clones⁹. Such mechanisms would act equally on synonymous and non-

synonymous substitutions and therefore predict a constant ratio of d_s/d_n during infection.

Phylogenetic analysis of the viral sequences reveals a division into two major lineages after year 3 (see ref. 4), which correspond to preferentially macrophage-tropic or T-cell-line-adapted phenotypes¹⁰. Comparison of the sequences in these two lineages (regardless to the sampling time point) reveals that the d_s/d_n ratio is significantly lower in the T-cell-line-adapted sequences than in the macrophage-tropic sequences (0.59 ± 0.37 and 1.39 ± 0.87 , respectively). This result was independently confirmed by analysing 177 additional V3 sequences from 133 different donors¹¹. For this dataset we found d_s/d_n ratios of 0.52 ± 0.16 for the T-cell-line-adapted sequences and 1.34 ± 0.45 for the macrophage-tropic sequences. This result can be explained if we assume that the immune response acts more strongly against the T-cell-line-adapted variant¹². In the context of the cell-tropism hypothesis, we would have to assume that T-cell-line-adapted variants evolve a larger variety of cell tropisms, which seems unlikely. Therefore, it is plausible that immune selection has a greater effect on V3 diversity than cell tropism (but the two factors may be interlinked¹³).

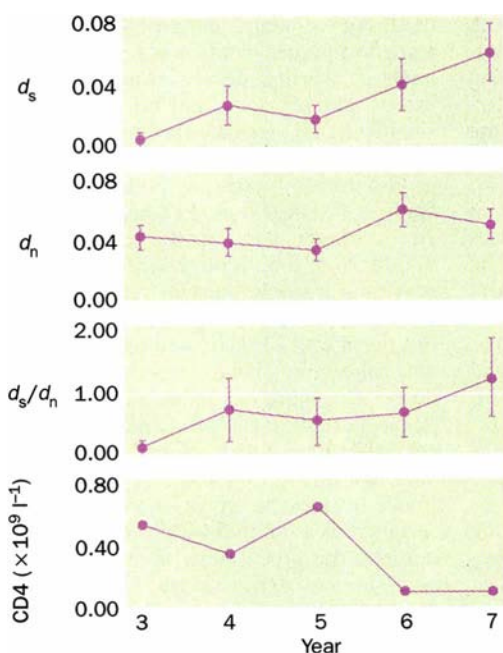
This is the only patient studied so far in sufficient detail to provide estimates of d_s/d_n ratios (with reasonably small standard deviations) of individual points. More data of this kind are urgently needed if we are to understand the evolution of HIV in individual infections and its consequences for pathogenesis⁶.

Sebastian Bonhoeffer
Edward C. Holmes
Martin A. Nowak
 Department of Zoology,
 University of Oxford,
 South Parks Road,
 Oxford OX1 3PS, UK

1. Ji, J. & Loeb, L. A. *Virology* **185**, 323–330 (1994).
2. Wei, X. et al. *Nature* **373**, 117–122 (1995).
3. Ho, D. D. et al. *Nature* **373**, 123–126 (1995).
4. Holmes, E. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4835–4839 (1992).
5. Zhang, L. Q. thesis, Univ. Edinburgh (1993).
6. Nowak, M. A. et al. *Science* **254**, 963–969 (1991).
7. Wolfs, T. F. et al. *Virology* **185**, 195–205 (1991).
8. Boyd, M. T. et al. *J. Virol.* **67**, 3649–3652 (1993).
9. Wain-Hobson, S. *Curr. Topics Microbiol. Immun.* **176**, 181–193 (1992).
10. Milich, L., Margolin, B. & Swanstrom, R. *J. Virol.* **67**, 5623–5634 (1993).
11. LaRosa, G. J. et al. *Science* **249**, 932–935 (1990).
12. Bou-Habib, D. et al. *J. Virol.* **68**, 6006–6013 (1994).
13. McKnight, A. et al. *J. Virol.* **69**, 3167–3170 (1995).
14. Nei, M. & Jin, L. *Molec. Biol. Evol.* **6**, 290–300 (1989).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.



Mean d_s , d_n and d_s/d_n for plasma-derived V3 sequences obtained from an infected haemophilic patient and the corresponding CD4 cell count at years 3, 4, 5, 6 and 7 after seroconversion. The d_s/d_n ratios indicate initially strong selection pressure for amino-acid change, which then declines over time. The viral DNA sequences (GenBank accession numbers M84240–M84317) were extracted using nested polymerase chain reaction and direct sequencing⁴. The standard deviations of d_s and d_n are computed according to ref. 14. An approximation for the standard deviation of d_s/d_n is given by the propagated error ($\Delta d_s/d_n + \Delta d_n/d_n$).

The Arctic's shrinking sea ice

SIR—A feature that has shown up in many general circulation model (GCM) simulations of greenhouse-gas-induced climate change is warming in the polar regions, particularly the Arctic¹, giving rise to a substantial retreat of the sea-ice cover². Observed reductions in the polar sea-ice cover could provide an early indication of greenhouse warming³. Microwave remote-sensing data from 1978–87 have revealed a slight but significant decrease in the extent of Arctic sea ice^{4,5}. We now find that subsequent microwave data suggest that over the period 1987–94 the rate of decrease in the extent of the Arctic sea ice has accelerated. It is too early to say, however, whether this represents a long-term trend.

The most direct, consistent source of information on sea-ice extent is microwave remote sensing from polar-orbiting satellites. The microwave data permit the retrieval of sea-ice concentration, from which ice extent (the area within the ice–ocean margin), ice area (the area of ice-covered ocean) and water area (the difference between ice extent and ice

	Arctic				Antarctic			
	Slope (10^6 km^2 yr^{-1})	S.E. ($\pm 10^6 \text{ km}^2$ yr^{-1})	% Change per decade	s	Slope (10^6 km^2 yr^{-1})	S.E. ($\pm 10^6 \text{ km}^2$ yr^{-1})	% Change per decade	s
Sea-ice extent								
1978–87	−0.032	0.008	−2.5	0.99	−0.018	0.017	−1.8	0.73
1987–94	−0.054	0.017	−4.3	0.99	+0.004	0.016	+0.4	0.20
Sea-ice area								
1978–87	−0.030	0.009	−2.8	0.99	−0.027	0.018	−2.7	0.86
1987–94	−0.049	0.016	−4.5	0.99	−0.003	0.016	−0.3	0.16
Water area								
1978–87	−0.001	0.006	−0.6	0.26	−0.009	0.008	−4.2	0.75
1987–94	−0.005	0.006	−3.2	0.60	+0.008	0.007	+3.8	0.70

Data for 1978–87 are from the Nimbus-7 SMMR and those for 1987–94 are from the DMSP SSM/I. The slope of each estimated least-squares linear regression line, standard error of the estimates (S.E.), the decadal trend expressed relative to the mean, and statistical significance (s) are shown. Bold trends are significant at the 95% confidence level or greater.

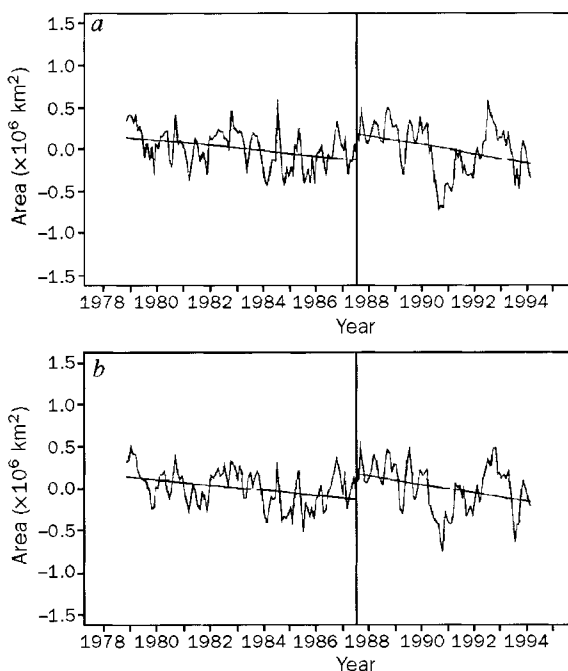
area) are derived. Previous analyses have emphasized the 8.8 years of Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) data from October 1978 to August 1987 (refs 4–6). Analysis using a band-limited regression technique revealed a statistically significant decrease in Arctic sea-ice extent of $0.032 \times 10^6 \text{ km}^2 \text{ yr}^{-1}$ (2.4% per decade), with no significant trends in the Antarctic⁵.

The follow-up sensor, the Special Sensor Microwave Imager (SSM/I) on board Defense Meteorological Satellite Program (DMSP) satellites, has provided data since

July 1987, including a 6-week period when both SMMR and SSM/I operated. In principle, the overlap period should permit inter-calibration between the sensors and a continuous, merged time series. At present, it remains prudent to treat the time series separately, because the differences in brightness temperature (T_B) calibration, sensor frequencies, viewing geometry and spatio-temporal coverage make inter-calibration difficult⁷. Time series of ice extent, ice area and water area were derived from ice concentrations, using the 15% ice concentration isopleth to define the ice edge, consistent with previous studies^{4,5}. The high-frequency variability was reduced by computing the mean in each month. Departures from the mean and from seasonal cycles were calculated using linear regression. These methods were applied to each hemisphere, although here we emphasize the Arctic.

The follow-up sensor, the Special Sensor Microwave Imager (SSM/I) on board Defense Meteorological Satellite Program (DMSP) satellites, has provided data since July 1987, including a 6-week period when both SMMR and SSM/I operated. In principle, the overlap period should permit inter-calibration between the sensors and a continuous, merged time series. At present, it remains prudent to treat the time series separately, because the differences in brightness temperature (T_B) calibration, sensor frequencies, viewing geometry and spatio-temporal coverage make inter-calibration difficult⁷. Time series of ice extent, ice area and water area were derived from ice concentrations, using the 15% ice concentration isopleth to define the ice edge, consistent with previous studies^{4,5}. The high-frequency variability was reduced by computing the mean in each month. Departures from the mean and from seasonal cycles were calculated using linear regression. These methods were applied to each hemisphere, although here we emphasize the Arctic.

The departures and significant trends for the Arctic, and the statistical



Departures in sea-ice extent (a) and sea-ice area (b) in the Arctic, 1978–94, derived from microwave radiative data from the Nimbus-7 SMMR (1978–87) and DMSP SSM/I (1987–94) data. The SSM/I data are 1-day polarized T_B s measured at 19, 22, 37 and 85 GHz. Ice concentrations were computed with the NORSEX algorithm¹⁰, which uses data at 18 and 37 GHz (SMMR) and 19 and 37 GHz (SSM/I). The statistically significant linear trends are indicated. The departures were calculated separately for the SMMR and SSM/I series, based on their different means. To perceive the trend for the full record, the 1987–94 departures should be shifted down by 0.29 and $0.22 \times 10^6 \text{ km}^2$ in a and b, respectively.

1. Stouffer, R. J., Manabe, S. & Bryan, K. *Nature* **342**, 660–662 (1989).
2. Manabe, S., Spelman, M. J. & Stouffer, R. J. *J. Clim.* **5**, 105–126 (1992).
3. Walsh, J. E. *Nature* **352**, 19–20 (1991).
4. Parkinson, C. L. & Cavalieri, D. J. *J. geophys. Res.* **94**, 14499–14523 (1989).
5. Gloersen, P. & Campbell, W. J. *Nature* **352**, 33–36 (1991).
6. Parkinson, C. L. *J. geophys. Res.* **97**, 14377–14388 (1992).
7. Zabel, I. H. H. & Jezek, K. C. *J. geophys. Res.* **99**, 10109–10120 (1994).
8. Cubasch, U. et al. *Clim. Dynam.* **11**, 71–84 (1995).
9. Hasseimann, K. *J. Clim.* **6**, 1957–1971 (1993).
10. Svendsen, E. et al. *J. geophys. Res.* **88**, 2781–2791 (1983).

cryospheric components, including snow cover, permafrost and ice sheets, together with temperature data, to help detect the fingerprints⁹ of greenhouse global warming.

Ola M. Johannessen*

Martin Miles

Einar Bjørge

Nansen Environmental and Remote Sensing Center,

Edvard Griegsvei 3a,

5037 Solheimsvik/Bergen, Norway

*Also at Geophysical Institute, University of Bergen, 5007 Bergen, Norway.

What do colour-blind people see?

SIR — Most human observers enjoy trichromatic vision: a three-dimensional colour space represents all the light that they can discriminate¹. Two per cent of the male population are dichromats, that is, they lack one class of photopigment and have a colour gamut that can be represented in a two-dimensional space. Here we offer to the normal trichromat a simulation of the reduced colour gamut seen by the dichromat. Our algorithm incorporates a strictly colorimetric transformation, but must also make explicit assumptions about the residual sensations experienced by dichromats.

For the normal trichromat, all discriminable light can be represented in three-dimensional space, with axes corresponding to the signals of the long-wave (L), middle-wave (M) and short-wave (S) cones²⁻⁵. Protanopes, deuteranopes and tritanopes lack L, M and S photopigments, respectively⁶. They confuse lights that differ only in the excitation of the missing class of cones; one of the three vector components is physiologically undetermined. In this simulation, we treat dichromatic vision as a reduction of the vision of the average normal observer, although both normal and dichromatic vision are known to be polymorphic⁷ and refinements of our algorithm could simulate alternative forms of, say, deuteranopia for alternative types of normal.

For the physiologically undetermined component, we choose values so as to imitate, for a normal observer, the appearance of colours for the dichromat. To this end, we make the following assumptions based on the colour perceptions reported by unilateral dichromats. First, we assume that neutral colours for normal observers are perceived as neutrals by dichromats. Accordingly, neutral stimuli, including any metamer, are not changed by the simulation and define the achromatic axis in the LMS colour space for both the dichromat and the normal. Second, we assume that the achromatic axis divides the surface of reduced



Reproduction of the video monitor display that simulates the reduced colour gamut of dichromatic colour-defective observers. *a*, Photograph of the 'Jardin des Plantes' (photo: Jean Le Rohellec; Grande Galerie, FNAC). The contrast was reduced to allow for all three projections onto the reduced stimuli surfaces to exist. *b*, *c*, *d*, Simulations of how *a* is seen by a protanope, deuteranope and tritanope, respectively. (A colorimetrically exact reproduction of the video display cannot be guaranteed in the printed version.)

stimuli into two half-planes, each corresponding to a single hue. Third, we assume that a stimulus of 575 nm is perceived as the same yellow, and a stimulus of 475 nm as the same blue, by trichromats as by protanopes or deuteranopes⁸. In a case of unilateral acquired tritanopia, the corresponding two hues were a red–pink at a dominant wavelength of 660 nm, and a blue–green at a dominant wavelength of about 485 nm (ref. 9).

Each half-plane of reduced stimuli in LMS space encompasses the achromatic axis and is anchored on a point specifying the invariant monochromatic radiation. Given the LMS specification of a stimulus, the algorithm replaces the undetermined component by the value corresponding to the projection of the original stimulus on the reduced stimuli surface, parallel to the direction of the missing fundamental axis.

The figure presents our simulation of the reduced colour gamut seen by each class of dichromat, after transformation of each pixel of a digitized picture. The L, M and S tri-stimulus values of each of the red, green and blue primary colours of the video monitor were determined by weighting their measured radiance spectra by the L, M and S sensitivities⁵. A protanope and a deuteranope were each satisfied with the match between the original image and the transformation corresponding to their particular deficiency, but they discriminated the original from the other transformations.

These responses of dichromats serve as a necessary check on our colorimetry; however, for any given image there are an

infinite number of alternative reductions that are metameric images for a particular type of dichromat. In using the rare cases of unilateral dichromacy to infer the residual sensations of the ordinary daltonian, we have followed an established tradition⁸. We do, however, recognize that the existence of unilateral dichromacy in male observers still lacks a genetic explanation and that the 'good' eye in such cases is seldom entirely normal¹⁰. There is one positive advantage in the particular wavelengths we have chosen to represent the residual sensations of protanopes and deuteranopes. At 575 nm, the Abney effect is minimized, that is, there is little change of hue at this dominant wavelength as the purity of the stimulus changes¹¹. At 475 nm, the loci of constant hue are curved, but mainly when they approach the spectrum locus¹². Within the colour gamut achievable on a video screen, stimuli within each half-plane of

1. Dartnall, H. J. A., Bowmaker, J. K. & Mollon, J. D. *Proc. R. Soc. B* **220**, 115–130 (1983).
2. MacLeod, D. I. A. & Boynton, R. M. *J. opt. Soc. Am.* **69**, 1183–1186 (1979).
3. Smith, V. C. & Pokorny, J. *Vis. Res.* **15**, 161–171 (1975).
4. Vos, J. J., Estévez, O. & Walraven, P. L. *Vis. Res.* **30**, 937–943 (1990).
5. Stockman, A., MacLeod, D. I. A. & Johnson, N. E. *J. opt. Soc. Am.* **A10**, 2491–2521 (1993).
6. Nathans, J., Plantani, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. *Science* **232**, 203–210 (1986).
7. Winderickx, J. et al. *Nature* **356**, 431–435 (1992).
8. Judd, D. B. *J. Res. natn. Bur. Stand. U.S.A.* **41**, 247–271 (1948).
9. Alpern, M., Kitahara, K. & Krantz, D. H. *J. Physiol., Lond.* **335**, 683–697 (1983).
10. de Vries-de Mol, E. C. & Went, L. N. *Clin. Genet.* **11**, 15–27 (1971).
11. Abney, W. de W. *Proc. R. Soc. B* **183**, 120–127 (1910).
12. Burns, S. A., Elsner, A. E., Pokorny, J. & Smith, V. C. *Vis. Res.* **24**, 479–489 (1984).

reduced stimuli look uniform in hue to a colour-normal observer, as they are assumed to look to a dichromat.

The availability of computer-controlled colour displays has allowed us to develop something that has often been asked for by nonspecialists — a simulation of how the dichromat perceives a complex coloured scene. Our algorithm should be of value to those who prepare display screens and colour-coding systems for use by the public. Although the quality of another's sensations can never fully be known, the present simulation illustrates, for the normal observer, the range of the daltonian's colour experience.

F. Viénot

H. Brettel*

Muséum National d'Histoire Naturelle,
CNRS UPR 257, 43 rue Cuvier,
75005 Paris, France

L. Ott

A. Ben M'Barek

Ecole Nationale Supérieure de
l'Electronique et de ses Applications,
95014 Cergy-Pontoise, France

J. D. Mollon

Department of Experimental Psychology,
University of Cambridge,
Cambridge CB2 3EB, UK

*Present address: CNRS URA 820, Département Images,
ENST, 46 rue Barrault, 75013 Paris, France.

How 'bad genes' survive

SIR — Wilcockson *et al.*¹ report the results of a study of genetic variance of body size in the seaweed fly (*Coelopa frigida*); large males are preferred by females in mating. Most of the genetic variance is expressed in males and is located in the $\alpha\beta$ heterotic-inversion system. The authors state, "Because heterotypic larvae exhibit superior viability, mate preferences based on size could generate fitter offspring, a slightly unusual type of 'good genes' female choice."

The fact that there is considerable additive genetic variance for size indicates that large size itself is not heterotic, and the authors do not say this. The suggestion of some kind of female choice in a heterotic system does, however, highlight a problem with the 'good genes' theory. The question is: what maintains the genetic variation from which the females choose? Why are 'bad genes' present to be rejected? Ewens² has shown that, at equilibrium, for a range of selection regimes there is no additive genetic variance for fitness. Charlesworth³ has discussed this in relation to the 'female choice' problem.

For the two-allele heterosis case, it is easy to show that, at equilibrium, with random mating the fitness of the offspring mated to the superior heterozygous males is the same as that of females mated to

FITNESS OF THE OFFSPRING OF FEMALES
MATED TO MALES BEARING ALLELES
A1 AND/OR A2

	A1A1 males	A1A2 males	A2A2 males
Female offspring	0.6775	0.6770	0.6740
Male offspring	0.6084	0.6101	0.6117

either one of the two inferior homozygotes and is equal to $1 - s/(s+t)$, where s and t are the selection coefficients against the two homozygotes relative to the heterozygote. If females choose to mate with the superior heterozygous males, then there is additional sex-limited heterosis in the males. This case cannot be solved analytically, but is amenable to simulations. For example, if the relative viabilities of the three genotypes in both sexes are 0.3, 1 and 0.4 and the relative mating successes of the male genotypes are 0.5, 1 and 0.7, the equilibrium frequencies of the three adult genotypes are 0.0938, 0.7339 and 0.1720. The table shows the mean fitness, at equilibrium, of the offspring of the females depending on the males to which they mate. The male offspring fitnesses include their future mating success and are lower because of the increased selection against the homozygotes relative to heterozygotes.

A survey of parameters involving more or less intense selection and symmetric as opposed to asymmetric selection on homozygotes gives the same result of essentially equal offspring fitnesses. The extremely slight differences in offspring fitness always show the offspring of preferred heterozygous males as being intermediate.

We emphasize that this is not meant to be a model of the seaweed fly system which is more complicated, including polygenic variation. We do propose, however, that the same problem applies when variation is maintained by simple balancing selection. The purpose of this simple exercise is to provide a reminder of a theoretical conundrum posed by the cause of variation in 'good genes' theories. Charlesworth³ and others have suggested solutions to this problem, such as cycling variation, migration, antagonistic pleiotropy and mutation selection equilibrium; these possibilities should be pursued when female choice is encountered.

Timothy Prout

Norman Eaton

Center for Population Biology,
University of California,
Davis, California 95616, USA

WILCOCKSON ET AL. REPLY — 'Good genes' sexual selection usually refers to the selection of mates with high genetic quality, and the subsequent production of progeny of high fitness. In spite of the general acceptance of good genes as an explanation for female mate choice (particularly outside the specialist genetic literature and in

popular television programmes), there are serious difficulties in understanding exactly how it could work as an evolutionary process³. The problem is that classical population-genetic theory predicts that no additive genetic variation of fitness should exist at equilibrium; and yet we have demonstrated high heritability for a sexually selected character, male size¹. This apparent contradiction is resolved if it is recognized that fitness is a theoretical concept of which experimental biologists can only measure components. In addition, Pomiankowski and Møller⁴ have recently provided a theoretical basis for understanding how genetic variability in sexually selected traits could be maintained.

We suggested that in seaweed flies there exists a "slightly unusual type of good genes female choice". If females prefer to mate, not with males of high viability as in the model of Prout and Eaton, but with the less viable homozygotes (strictly speaking, homokaryotypes), and if females mate disassortatively, then the progeny produced are indeed fitter⁵. In heterotic systems, homozygous females mating with males homozygous for a different allele produce heterozygous offspring with the highest viability. The mating preferences of heterozygous females are of little consequence in terms of offspring viability. Observations on some, though not all, natural populations of seaweed flies suggest that females do exhibit such a pattern of mating⁶. This means that males cannot be said to have 'good' or 'bad' genes; it is the complementarity of sperm and eggs that is important.

In seaweed flies, beauty is in the eye of the beholder, and is judged on the basis of size. Although size itself is not heterotic, its major genetic determinant, a polymorphic chromosomal inversion system, is. Furthermore, the genes determining female preference may well be in linkage disequilibrium with this same inversion system^{7,8}. It will be interesting to see if heterotic systems in other species are subject to 'good genes' sexual selection of this sort. Perhaps we shall be proved wrong in saying that mate choice in female seaweed flies is of an unusual type.

Ray Wilcockson

Caroline Crean

Tom Day

Department of Genetics,
University of Nottingham,
Nottingham NG7 2UH, UK

1. Wilcockson, R. W., Crean, C. S. & Day, T. H. *Nature* **374**, 158–159 (1995).
2. Ewens, W. J. *Genetics* **83**, 601–607 (1976).
3. Charlesworth, B. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 21–40 (Wiley, Chichester, 1987).
4. Pomiankowski, A. & Møller, A. P. *Proc. R. Soc. B* **260**, 21–29 (1995).
5. Gilburn, A. S. & Day, T. H. *Heredity* (in the press).
6. Gilburn, A. S. & Day, T. H. *Proc. R. Soc. B* **255**, 159–165 (1994).
7. Gilburn, A. S., Foster, S. P. & Day, T. H. *Evolution* **47**, 1788–1795 (1993).
8. Gilburn, A. S. & Day, T. H. *Genet. Res.* **64**, 19–25 (1994).

Pixelating the populace

Lewis M. Branscomb

Being Digital. By Nicholas Negroponte. Knopf/Hodder and Stoughton: 1995. Pp. 243. \$23, £12.99.

READERS of *Wired* magazine are familiar with Nicholas Negroponte's columns, from which the best parts of this book are drawn. Leonardo da Vinci recognized that reality is most effectively described through the eye. Negroponte and his colleagues at the Media Lab at the Massachusetts Institute of Technology would agree, but would include the mind's eye (virtual reality) too. In their view, information should be experienced through all the senses and through the imagination.

Current debate about 'information infrastructure' concerns the providers of content (publishers and radio, television and movie producers) and conduit (telephone, cable, satellite and computer-network servicers). Negroponte sees conduit and content as inextricably linked, preferring to distinguish between the merits of digital versus analog information and services.

In the digital world, information is understandable to computers, which can quickly transform it for use by humans, trapped as we are in our hopelessly analog condition. But computers are still novelties to most people. The overwhelming majority of every country's information infrastructure today is comprised of the real, analog world of telephones and television. Only a third of homes in the United States have a computer, whereas about 93 per cent have a telephone and 98 per cent have a television set. Neither the telephone nor the television is now digital.

Analog information services are doomed to obsolescence in Negroponte's world. To sharpen the distinction between analog and digital information, he identifies analog information with the physical media through which it is carried, transformed and embodied. He calls physical media "atoms". Digital information is encoded in units of information called "bits". Negroponte celebrates the abstract nature of bits, and of bits about bits (the message headers that tell packet-switched networks what to do with the content). It is not enough, however, to be digital: we must embrace the multi-sensory world in which pixels (comprised of bits) are the elements of images. *Being Pixelated* as a title would have captured this thought and Negroponte's somewhat flippant but readable discourse.

Being Digital is written for the lay reader, not for the scientist or sophisticated computer user. The book's jacket commends it to those "lost on the

information highway", a misleading metaphor which Negroponte, to his credit, rarely uses. It claims, astonishingly, that this is the "first book to explore the impact of digital technology on the world", but more scholarly books on the subject have appeared by the hundreds since the late 1970s. The text is divided into three sections. The first part is intended to help naive readers understand the other parts, although in fact it will probably confuse most lay readers and surely irritate the technically trained. Much of it is written in 'bitspeak' — the text equivalent of soundbites. Some statements remind one of Marshall McLuhan: "flight simulation... is more realistic than flying a real plane".

The most interesting part of the book is found in the second and third sections. Here Negroponte predicts the convergence of information media: television will begin to look more like a newspaper, offering a broad menu of information from which to choose; our personal computer will make each of us our own publisher, if we let it know enough about our individual needs; and our personal "interface agent" will browse through the huge choice of multimedia information and select just what we want, packaged in the format and media we want.

His model of the right computer design is the English butler. In the "post information age" foreseen in this book, everything

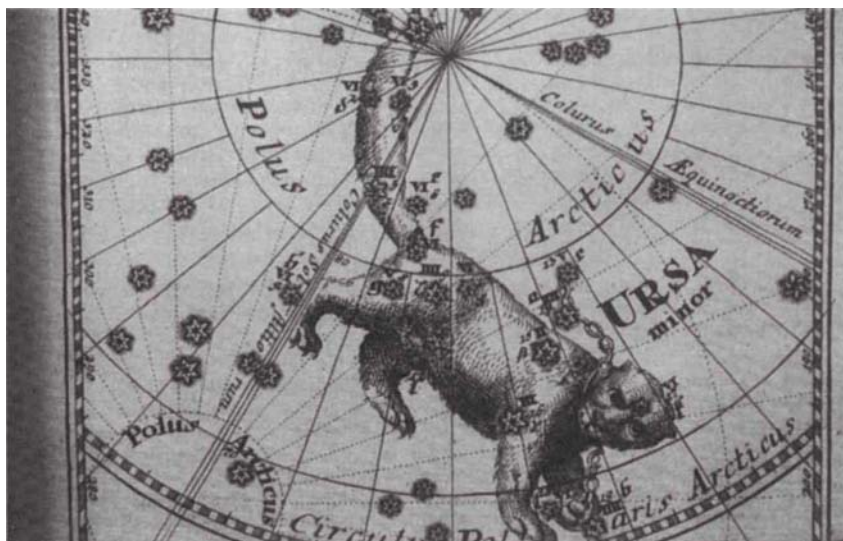
is individualized: the audience size is one; being digital means I am unique.

'Multi-media' must be viewed as an integrated sensory experience to be effective. When Russ Neuman showed Media Lab focus groups identical television signals with different audio qualities, observers selected those signals with the best sound quality as having a superior picture — which is why Negroponte wears his glasses when he eats.

The least satisfactory part of the book is its view of how law and economics will influence the realization of Negroponte's vision. The author suggests that copyright law will "break down completely before it is corrected" — even John Perry Barlow, a critic of copyright and a regular contributor to *Wired*, doesn't take such an extreme view. And he worries that readers of an electronic *New York Times* will bypass it and subscribe directly to their favourite writers through the Internet. The reason this doesn't happen today is that the *New York Times* knows how to write an employment contract.

Survivors of the publishing industry will transform themselves into trusted information brokers and suppliers of information agent software. As Negroponte points out, the value of information about information can exceed the value of the information itself. Its ultimate value depends on reliability, comprehensibility and suitability for use. People smarter or better informed than ourselves will always be paid to be the trusted editors of good books or the authors of trusted agent software. □

Lewis M. Branscomb is at the John F. Kennedy School of Government, Harvard University, 79 J. F. K. Street, Cambridge, Massachusetts 02138, USA.



Ursa Minor, from Corbinianus Thomas's *Mercurii Philosophici Firmanentum Firmianum* (1730). The picture appears in *The Cambridge Guide to the Constellations* by Michael E. Bakich. Cambridge University Press, £35, \$49.95 (hbk), £14.95, \$19.95 (pbk).

Brute forces

Roy Porter

A History of Medicine in the Early US Navy. By Harold D. Langley. *Johns Hopkins University Press: 1995. Pp. 435. \$49.95, £41.50.*

A REMARKABLE labour of love, Harold Langley's substantial volume records the lives of early US naval surgeons, the engagements in which they were involved and the casualties they treated, in painstaking and often gory detail. During the war of 1812 against the United Kingdom, Amos Evans was surgeon on the *Constitution* which grappled with HMS *Java* off the South American coast. Seaman Peter Furanse was injured. He had been severely wounded by grapeshot that lodged in his Achilles tendon. The ball was removed, the leg splinted and hot poultices applied. Furanse was purged and an "antiflogistic [*sic*] regimen" attempted. He became a distressing post-operative case. After two days he discharged a great amount of pus from his wound; eight days later he died.

A couple of hundred pages of Langley's book are given over to particulars such as these. At times it seems as if every case of yellow fever or venereal disease (known as 'ladies fever', for which the surgeon expected private payment) and every tumble from a mast gets its mention. All this makes it a rather gruelling read, however valuable for the scholar as documentation.

Such wounds, treatments and deaths were of course the bread-and-butter of

naval surgery in every fighting nation. So too was a certain brutality. One of the very first surgeons employed by the US Navy, Charles Webb, was cashiered for sticking his dirk into the chest of a black cabin boy and flogging a seaman. (Barbarity of that order was at least noted and punished.) Edward Cutbush, a career naval surgeon and the hero of Langley's story, was to complain that the service was dogged by an unwholesome image: "I can assure you that the description of a naval surgeon, by the pen of the celebrated Dr Smollett in his *Roderick Random*, has prevented many men of professional abilities from entering our service, under an idea that the surgeons and mates were considered in the same menial situation". But in truth it hardly needed novel-reading to convince young medical men that naval service was neither well remunerated nor well respected.

US naval surgeons typically lacked a university degree or an infirmary training. Unlike their colleagues in Europe, they had not attended Hunter's anatomy school or the Paris hospital; they had picked up their skills by apprenticeship and had then been recruited haphazardly. Aboard ship, they therefore occupied a highly equivocal place on the social ladder, often being treated with blatant disdain by the officer class. Surgeon Shannon was drinking one evening with the officers on the brig *Scammel*. They fell to debating the meaning of the olive branch on the Great Seal of the United States. The surgeon ventured his view, whereupon he was informed by Lieutenant Ludlow in no uncertain terms that he had no right to an opinion on the subject. An argument broke out, pistols were

fetched, there was talk of duels; and the incident ended with Shannon's resignation from the service.

None of these rather shambolic touches is surprising, as the US Navy was new — it was established only in 1794, with the commissioning of four frigates — and somewhat limited in its operations, being viewed primarily as an instrument of coastal defence rather than, as with Britain or Napoleonic France, an arm of imperial policy. Naval surgery therefore led a somewhat hand-to-mouth existence. There was long-running congressional resistance to setting up Navy hospitals. The fear was of jobbery and speculation, a worry amply justified when, for instance, Benjamin Waterhouse, the surgeon in charge of the Boston hospital, was caught trying to get rich quick by salting away stores and funds by the tactic (all too familiar nowadays from the sleazy actions of the boards of our own privatized utilities) of assigning them to his wife, under her maiden name. In place of naval hospitals, sick sailors, where possible, were off-loaded into civilian hospitals ashore. The United States developed no big naval medical establishments and, partly by consequence, no valiant figures of the stature of James Lind, Thomas Trotter or Sir Gilbert Blane.

In such circumstances, it is no wonder that the first US book of navy medicine, Cutbush's *Observations on the Means of Preserving the Health of Soldiers and Sailors* (1808), was essentially cribbed from the pioneering writings of Trotter. Like the British naval physician, Cutbush stressed the need for hygiene and ventilation, the importance of warmth aboard ship during the winter and the havoc that habitual intoxication could wreak on sailors' health.

Yet it would be wrong to paint too negative a picture. A few individuals campaigned long and hard for improvements, especially Cutbush, who had studied medicine in the 1790s at Pennsylvania Hospital and been appointed as surgeon to the frigate *United States*. In due course, naval hospitals were set up in Boston, New Orleans and so on, with an asylum in Philadelphia. From 1823, lectures were given, especially to assistant surgeons, to improve their skills; entrance examinations were later introduced; and as a culmination of the growth of professionalism the Bureau of Medicine and Surgery was established in 1842, by which time the Navy Department had on its roll 61 surgeons and around the same number of assistants. At least from that time onwards, the charge that the service was riddled with Smollettian savagery would no longer stick. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.



Anti-venereal-disease poster designed for use among the Allied troops during the First World War. Both the incidence and the fear of syphilis dramatically increased until after the Second World War, when antibiotics began to be widely used. The poster appears as part of a small exhibition at the Wellcome Institute for the History of Medicine in London that runs until 30 September. Entitled "Fatal Attractions", it looks at AIDS and syphilis from medical, personal and public perspectives.

Being and becoming

H. M. Collins

The Construction of Social Reality. By John R. Searle. Free Press/Allen Lane: 1995. Pp. 241. \$25, £20.

JOHN Searle is a philosopher at the University of California at Berkeley. In 1984 he delivered the BBC Reith lectures, cementing his reputation as one of the clearest and most forceful thinkers around. He is probably best known outside the professional heartland for his 'Chinese room' argument against artificial intelligence, which he repeated in his Reith lectures. Searle imagined a person in a room managing conversational interchanges in Chinese by manipulating Chinese word-symbols according to a rule book; the person in the room might or might not understand Chinese. Thus Searle proved to nearly everyone's satisfaction that symbol manipulation is not the same as understanding.

In his latest book, Searle looks at the social sciences. The title is provocative in that it contrasts with that of a famous book by Peter Berger and Thomas Luckmann, *The Social Construction of Reality* (1967). Searle aims to show the difference between what can and what cannot be socially constructed. On the way he develops a refreshingly clear exposition of the problems of the social sciences, in the positive sense of problems that are difficult and interesting.

Searle explains that the social sciences, as opposed to the natural sciences, have to deal with things that exist only because we think they exist. Take paper money: on the one hand there is the actual paper and printing; on the other hand there is the value that resides in money only so long as everyone continues to believe, act and talk as though it is valuable. (Searle includes an excellent discussion of the role of language in the creation of 'institutional facts'.) Once people stop thinking, talking and acting collectively as though money is valuable, it stops being valuable. This is a philosophical puzzle because money is at least at real in its effects as subatomic particles — as the frustrated builders of the Superconducting Super Collider know. Searle thinks, then, that there is social construction of social things and that these things are nevertheless real.

Once one sees that things that exist only because we think they exist affect all our lives in a way that is as concrete as can be, the recent arguments between natural and social scientists are put into context. Social scientists are surprised that natural scientists have difficulty with this kind of idea. For example, Richard Dawkins

insists that there are no social constructivist at 30,000 feet who aren't hypocrites, yet if he has money in his pocket he is a social constructivist himself.

Where Searle differs from what he perceives to be the view of social constructivists is that he thinks the existence of social things presupposes a class of things that are there whether we think about them or not. I leave the details of the argument to the reader. Agree with him or not, in putting the matter so clearly, Searle shows the way to the interesting questions. Is it true that social things are based on nonsocial things? If it is true, where is the boundary between social and nonsocial? How do we tell where the boundary is? What constraints do nonsocial things place on the construction of social things and vice versa? If some social scientists have overstepped the boundary — and this may be the cause of the heat in recent debates — how can we argue the matter sensibly? How can we investigate the way in which facts come into being without each side simply trying to impose its authority? □

Harry Collins is at the Science Studies Centre, University of Bath, Bath BA2 7AY, UK.

Science evolving

Ray Percival

Evolutionary Naturalism. By Michael Ruse. Routledge: 1995. Pp. 316. £35, \$49.95.

MICHAEL Ruse aims to describe what scientists actually do in their research and how they arrive at their theories — a mixed bag of false starts, fallacious reasoning, the cultivation of followers, the marketing of ideas and so on. His approach, evolutionary naturalism, rejects the traditional distinction between the normative and the descriptive analysis of science. For him the path of discovery to, say, Darwin's theory of natural selection makes a difference to the theory itself, whereas for the normative analyst it is just history. Normative analysts (who probably include most readers of *Nature*) would say that the logical structure of the theory, its truth or falsity and its relevance to the objective problem can all be assessed independently of the route of discovery.

A scientist's problem is to produce an explanatory theory of greater truth and depth than any rival theory; a look at the path of discovery might give us hints about how to interpret this objective problem situation. But, having said that, it is important to distinguish between Kekulé's tail-biting snakes and his problem situation (how to explain benzene phenomena), a distinction one

wishes Ruse had explored systematically.

It is worth stressing that problems, which Ruse (following the philosopher Larry Laudan) believes introduce an obviously subjective element, can be treated as objective abstract entities. Anyone who doubts this can consult the surprisingly interesting *Guidelines for Examination in the European Patent Office* (European Patent Office, 1994). It is clear from this document that a person's subjective conception of an objective problem may be wrong and may fail to be decisive in the eventual solution. Ruse writes as if Karl Popper never said a word about the evolution of scientific theories from objective problems.

Ruse's use of Thomas Kuhn to undermine Popper's falsification theory is a weak assault on normative analysis. In the *Structure of Scientific Revolutions* (University of Chicago Press, 1962), Kuhn says that "[no] process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature". This passage is the start of the myth that Popper was a naive falsificationist (that is, someone who believes in conclusive falsification) and of the confusion that his falsificationism was an historical thesis. Falsificationism was meant as a normative proposal based on a logical analysis of the situation facing the scientist eager to learn from mistakes made in blindly groping for the truth. Only secondarily was it meant to suggest what actually happens in science. Nevertheless, there are many interesting examples that conform to the pattern of Popper's conjecture and refutation, for example Rutherford's refutation of J. J. Thomson's theory of the atom in 1911. (For more, see Popper's *Realism and the Aim of Science*, Hutchison, 1983).

There are two strange things about the above passage from Kuhn. First, we are supposed to regard it as a falsification of falsificationism. But why should we, if, as Ruse insists, scientists ignore falsifications? The naturalist does not have an answer, because he cannot tell you what you should or should not do. Second, rhetorically the argument trades on the tacit assumption that scientists mostly get things mostly right (and if there is a best method, then they will be using this soon if not now). But, being fallible, all of them may one day get it not just mostly, but completely wrong (or at least overlook a better method). And in fact, they have. The naturalist defines away this possibility. The normative analyst can also ask: how can we promote the growth of scientific knowledge? What method(s) should the scientist adopt if this is his aim? How should we control error? All these questions are lost in naturalism.

Ruse does shy away from a crude scientism that says that all problems can be

solved by science, but there are many forms of naturalism and it is unclear exactly where Ruse stands (see Susan Haack's *Evidence and Enquiry* (Blackwell, 1993) for clear distinctions and refutations of some forms of naturalism).

In the section on evolutionary epistemology, Ruse argues that scientific thinking (indeed, all thinking) is to be understood with the help of "epigenetic rules". These are rules of thinking that we are disposed to follow because of our evolutionary history, such as the law of non-contradiction. If our ancestors were inclined to regard sabre-toothed tigers as both dangerous and not dangerous at the same time, they would not have been our ancestors. The approach is much more promising than that of, say, the philosopher David Hull, who sees the scientific success of a theory simply in terms of the proliferation of memetic copies, without considering the rational filters at work. But, although promising, Ruse's approach shows its defects when he goes on to say that the necessity of logical rules is based on nothing but our feelings of necessity shaped by evolution. He neglects the fact that conforming to the law of non-contradiction saves us from moving in an argument from true premises to a false conclusion, and also enables us to test the implications of our theories (because a false conclusion of a valid argument implies a false premise). Logical necessity and validity are emergent properties of language and cannot be defined in terms of feelings (evolved or not).

Such a view also neglects the most interesting thing about evolution — the emergence of radically new structures and properties not fully derivable from, contained in or explicable by those that came before. Once we have language, for instance, we can build theories that no one person can fully grasp: they are no longer just a part of our psychology. Perhaps the ability to count, for example, was selected phylogenetically, but after Godel's famous incompleteness proof, we must say that the linguistically formulated theory of natural numbers has unfathomable content and that this is a fact independent of the way we feel about the proof. If this is not enough to convince Ruse that we are capable of producing systems that have independent properties, I would suggest looking at Danny Hillis's work, which suggests that self-evolving computer programs may produce perfectly working airline navigation systems containing millions of lines of program that nobody could possibly understand.

There are surely elements of marketing in the selection of scientific theories, but, if one wants to know about that, it is wise to consult advertising and marketing specialists. At the end of the day we have still to analyse what is the best method for enhancing the growth of scientific truth.

We can teach students how to falsify theories and how to obtain jobs, and with enough ingenuity, there need be no conflict.

I have focused mainly on a few problems in *Evolutionary Naturalism*, but it covers in an informative way many other interesting areas. In the discussion of evolutionary ethics, for example, Ruse argues that our morals are, like our logic, based on biological evolution. The book is a readable and forthright presentation of his views on the relevance of evolution to philosophy and could serve as a useful introduction for students. □

Ray Percival, an associate editor of *Journal of Social and Evolutionary Systems*, is at 70 Hillview Court, Astley Bridge, Bolton BL1 8NU, UK.

Relative values

Sean Nee

Macroecology. By James H. Brown. University of Chicago Press: 1995. Pp. 269. \$42.50, £33.95 (hbk); \$15.95, £12.75 (pbk).

FROM 1736, generations of the Marham family, who lived near Norwich, England, each year meticulously recorded facts such as the dates of the first snowdrops and cuckoos, resulting in an invaluable data set for studying how species respond to variations in climate. Darwin provides the best example of how endless catalogues of facts about natural history, each fairly dull in itself, can be illuminating when looked at, imaginatively, as a whole.

This stamp-collecting instinct is rife in ecology. Starlings weigh 80 grams and four million of them live in Sweden. Ecologists do not of course stare blankly at these facts, mesmerized. Rather, they ask how and why starling numbers have changed over time, and how a starling's weight affects its chances of surviving the winter. These questions take us in one direction, towards the particular. James Brown invites us to go in the other direction and to ask, for example, about the relationship between body size and abundance of birds in general — that is, to put all the facts into one big data set, to find empirical generalizations and to try to understand them.

Clearly, there are many permutations. If we just look at the shape of the distribution of bird body sizes, we see that it is skewed, with many more small species than medium or large ones. This is true not only for birds but for mammals, fish, trees, bacteria and insects as well, suggesting that there is a general principle to be unearthed. The main variables for which macroecologists have data are the body

size, abundance, habitat specificity and geographical range of species, providing us with spaces of up to four dimensions in which to plot species points and look for interesting features in the resulting cloud. Brown demonstrates that with imaginative use of these data, meaningful hypotheses can be posed and many of their implications can be tested.

This is not a "radical new research agenda", as the publishers would have us believe, although it is interesting and important, and this is the first book to package the field for wide consumption. But it does contain some radical ideas. For example, Brown offers an argument to explain the distribution of animal body sizes that defines fitness as the rate at which animals invest energy in reproduction. Because this invites us to discard current theories of life-history evolution, the author indeed "makes a major intellectual leap", but perhaps into the abyss.

Our thinking about allometry is also challenged. It is well known that larger mammals live longer, bear fewer young, have higher metabolic rates and so on. In other words, the relationship between body size and just about everything else is mono-tonic. Brown suggests that many of these relationships may in fact be triangular, with very small-bodied species showing the opposite overall trend. If true, this would be important. But so far there is only anecdotal evidence, with one exception. Certain data sets reveal a triangular relationship between body size and abundance, with the most abundant species having intermediate sizes. But it has been argued in the literature that the relationship is merely a sampling artefact, although Brown does not mention this.

In fact, there is much that Brown does not discuss, and other macroecologists reading this book will, quite rightly, be annoyed by the extent to which their work is ignored. One is given no hint, for example, that people have been using species-area relationships to predict extinctions for decades before the publication of a paper coauthored by Brown in 1992, which gets a chapter all to itself. Nevertheless, Brown is a dominant figure in this field and his work is a suitable vantage point for an overview of the macroecological research programme. □

Sean Nee is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

■The *Chambers Dictionary of Science and Technology* is 55 years old this year and was last revised seven years ago. The seventh edition, edited by Peter Walker, has just been published as the *Larousse Dictionary of Science and Technology*. Its 1,236 pages contain 49,000 entries, over 500 illustrations and many appendices. Larousse, £45 (hbk), £19.99 (pbk).

Evolutionary theory of parent-offspring conflict

H. C. J. Godfray

Natural selection can act in different ways on genes expressed in parents and their young, giving rise to parent-offspring conflict. The way in which this genetic conflict manifests itself at the behavioural level is unclear, and there has been widespread dissatisfaction that the theory has provided few testable predictions. A recent shift in emphasis from models that define the possible extent of conflict to those that predict its resolution offers greater hope of a testable theory of parent-offspring conflict.

THE length and extent of parental care in animals that look after their young are determined by a complex set of behavioural interactions between parent and offspring¹. Since the pioneering work of Hamilton² and Trivers³, it has been widely appreciated that natural selection often acts in subtly different ways on genes expressed in the parent that influence parental care and on the equivalent genes expressed in the young. Although to a large degree the two sets of genes share a common evolutionary interest, parent-offspring conflict is predicted to occur over the amount and duration of parental care. Since its formulation, the subject of parent-offspring conflict has been one of the most contentious areas of behavioural and evolutionary ecology^{1,4}, and one that has proved among the most refractory to experimental investigation¹. Several authors have expressed dissatisfaction with the theoretical underpinning of the subject and even wondered whether parent-offspring conflict, while theoretically possible, is actually very rare in nature^{5,6}. However, our understanding of how natural selection moulds the relationship between parent and offspring has improved in recent years. Although it would be too optimistic to say that a perfect meshing together of theory and experimental studies is imminent, there are new hypotheses to test and a clearer picture of where new theory is required. Parent-offspring conflict theory has been applied to a variety of problems including interactions between social and parasitic insects^{4,7-12}, conflict over the sex ratio^{3,13,14}, the provisioning of seeds and other aspects of plant life history^{15,16}, helpers at the nest¹⁷ and biochemical interactions within the womb¹⁸⁻²⁰. In what follows, I shall largely concentrate on the classical problem, first addressed by Trivers, of the provision of food and other resources by avian and mammalian parents.

Parent-offspring conflict is easiest to understand from a kin-selection perspective, and easiest to explain using the anthropomorphic shorthand that so enrages some biologists. A gene expressed in the parent which influences the distribution of parental resources is equally likely to be present (identically by descent) in any of its offspring. All else being equal (that is, assuming no differences in offspring need and future reproductive value), the parent should give an equal share of resources to all its young. A gene expressed in the offspring which influences the distribution of parental resources will be selected to try to obtain a disproportionate share of resources. This is because the gene is certainly present in the focal offspring but present (again identically by descent) in any one of the offspring's siblings with a probability given by the coefficient of relatedness². In a brood of full siblings the coefficient of relatedness in birds and mammals is normally one-half; in broods of half-siblings the relatedness is one-quarter. Although

the offspring is selected to try to grab a larger share of resources for itself than the parent would wish, there is a limit to its selfishness as it still has a genetic interest in its siblings. In the simplest cases, whether natural selection favours an act of sibling selfishness is determined by Hamilton's rule: $b > rc$. The benefits (b) to the actor must outweigh the costs (c) to the victim weighted by the coefficient of relatedness (r).

The first full discussion of parent-offspring conflict by Trivers³ attracted much criticism⁴, some of which continues today. A persistent argument has been that parent-offspring conflict is logically flawed: any advantage gained when young is counteracted by the disadvantages of producing conflicting offspring when adult. This fallacy disappears if we move from an individual to a gene perspective²¹. Competition occurs each generation between conflicting and non-conflicting alleles at the same locus; the conflictor wins because in each generation individuals carrying the conflictor allele tend to obtain the more resources²¹⁻²⁴.

Much of the dissatisfaction with the theory of parent-offspring conflict stems from the plethora of models that exist in the literature, most technically very complex and often having very different goals. One class, battleground models²⁵, seek to calculate the optimum levels of resource distribution for genes expressed in parents and young. Such models are useful in exploring the potential for parent-offspring conflict but do not themselves make predictions about the outcome of parent-offspring conflict. For such predictions we require resolution models, which incorporate assumptions about the phenotypic interaction between the parent and offspring. These are more likely to provide testable predictions, although they also tend to be much harder to formulate and solve. Both types are concerned with the outcome of evolution or evolutionarily stable strategies (ESSs); other models investigate evolutionary trajectories, how gene frequencies change under the influence of natural selection acting on parents and young. Such models can be valuable in clarifying whether an ESS can be achieved by a population, but it seems unlikely, at least with present experimental models, that they will provide testable predictions. Finally, models can be classified by whether they use the technology of kin selection or explicit population genetics. Kin-selection models are undoubtedly the most powerful, in the sense that they can be used to analyse more realistic problems, but they require assumptions to be made about the strength of selection and the additivity of fitness effects. Where evolutionary endpoints rather than trajectories are studied, the two approaches tend to give the same result, although the study of some traits may require explicit genetic modelling. Given the immense difficulty in understanding the genetics underlying parent-offspring conflict, as well as the

problem in testing hypotheses about evolutionary trajectories, kin-selection resolution models seem to provide our best prospect of a testable theory of parent-offspring conflict.

Resolution models

Trivers' first major critic, Alexander⁴, argued that, even if parent-offspring conflict occurred (which at the time he doubted), the parent would always win as it was in the position to implement its chosen resource share through *force majeure*. Although a battleground existed, the resolution was always at the parental optimum. This is an example of perhaps the simplest resolution model where only one side is able to influence the outcome. The best examples of this type of result probably come from species without parental care. Consider an insect that lays its eggs in clutches and whose offspring compete for a limited resource. Assume further that the young can consume resources either fast and inefficiently, or slowly with greater efficiency. It is straightforward to construct battleground models showing that there is potential conflict over clutch size and over the manner in which the resource is consumed (competition among siblings selects for resource pre-emption and inefficient resource utilization)^{26,27}. However, the young cannot directly influence clutch size, and the parents are unable to regulate competition among their offspring. The resolution of each conflict is thus trivially determined by the party that is able to influence events. Nevertheless, more complicated outcomes are sometimes possible. If the insect larvae are cannibalistic they can influence brood size, if not clutch size, leading to an evolutionary game in which parents may lay excess eggs in the knowledge that brood reduction occurs^{4,28}. Alternatively, a parent may be selected to reduce clutch size if this results in less sibling competition, increased efficiency in resource utilization, and an overall rise in total brood fitness²⁶.

The more interesting situation is where there is the opportunity for both parents and young to influence resource allocation. Constructing a resolution model is now much more complex, and requires specific assumptions about the freedom of action of both parties. In the context of parental care in birds and mammals, the resolution model must explain resource allocation, but also the observation that there appear to be areas of conflict within family life. Offspring frequently solicit resources from their parents in ways that appear wasteful in terms of energy use, or that risk attracting predators^{1,29,30}. Furthermore, young often seem to throw tantrums, and parents and offspring tussle and squabble^{1,3}. There are clear dangers in overinterpreting such behaviours (especially as they are observed through the prism of one's own family experiences). Indeed, commentators have suggested that the use of the word conflict to denote both genetic conflicts of interests as well as observed squabbles within the family has led to the uncritical acceptance of the theory of parent-offspring conflict⁵.

Trivers³ put forward a verbal model of the resolution of parent-offspring conflict. He argued that the parent would be selected to monitor the requirements of its offspring and so be subject to 'psychological manipulation'. The young would attempt to misrepresent their needs, and perhaps age, so that they are fed more and so that the period of parental care is extended. Weaning tantrums and noisy solicitation are manifestations of the young's attempts at psychological manipulation. A different verbal resolution model was suggested by Zahavi³¹. Young engage, or threaten to engage, in behaviour that reduces their fitness, and the parent provides more resources to try to prevent this. In effect, the offspring blackmail their parents into providing more resources. While psychological manipulation and blackmail are potent metaphors, can they really arise through natural selection operating on interactions within the family? Quantitative models are essential to answer this question.

The first quantitative battleground and resolution models were constructed by Parker & Macnair^{32,35}, who provided an explicit genetic model of parent-offspring conflict at a time when it was widely thought to be a population-genetic impossibility (see also

refs 36–38). Their resolution model suggested that the ESS resource distribution would be intermediate between that of the parent and offspring, a result that has been widely quoted as the most likely outcome of parent-offspring conflict. In the last few years a number of new resolution models have been developed which incorporate different aspects of both Trivers' and Zahavi's ideas, and that make varying predictions about the ESS resource distribution. Unfortunately, comparison between the models is immensely difficult because they are phrased in different ways (as phenotypic or genetic models, predicting ESSs or evolutionary trajectories), and because they make radically different assumptions about the mechanisms of parent-offspring interactions. I have attempted to extract the underlying biological mechanisms behind the four main resolution models and to incorporate them in a single setting where their similarities and differences can be seen more easily (Box 1). I consider the case of a (single) parent that rears one offspring a year. I assume a negative correlation between the resources expended on the current young and the parent's future reproductive success. The period of parental care consists of a series of episodes that have independent effects on offspring fitness. Clearly, many of these assumptions will have to be relaxed before specific predictions can be made about individual systems (see below).

The four resolution mechanisms described in Box 1 fall into two classes. In the first, the resolution is determined by a balance between parental and offspring behaviour: deviations from ESS behaviour by either party elicit a reaction in the other that negates any advantage. These mechanisms seem to be in the spirit of Zahavi's verbal model. In the second type of resolution mechanism, parental behaviour is determined by offspring behaviour, but not vice versa (except indirectly through the resources the young receives). For the resolution to be an ESS, the offspring's behaviour must encode fundamentally reliable information for the parent, although there is scope for some misrepresentation as envisaged by Trivers' resolution argument.

The Parker-Macnair model is of the first type: the parent responds to offspring behaviour (e.g. begging) by providing more food, and the offspring responds to increased resources by begging less. At the ESS, the parent provides more food than the parental optimum in the absence of solicitation to prevent the offspring from further lowering its fitness by expensive begging. One possible problem with this model is that the parental response to offspring begging and the offspring response to more resources are assumed and not derived. It has yet to be shown that the patterns of behaviour that allow blackmail to evolve can themselves be evolutionarily stable.

Yamamura and Higashi³⁹ assume that both parent and young can influence resource share, but pay a cost in trying to manipulate the outcome. Any deviation from the ESS by either parent or offspring causes a conflict that entails costs for both parties. The exact value of the ESS resolution depends on the relative costs that the parent and young experience when attempting to determine resource level. This is a general model of the resolution of conflict between relatives that has not yet been applied specifically to parent-offspring conflict. There is a need for the assumptions it makes about the nature of the behavioural interaction between parent and young to be assessed within the framework of a more mechanistic model of parent-offspring conflict.

In the other two models described in Box 1, the parent monitors offspring behaviour and uses the information it gains to implement its optimum resource allocation. The mechanism put forward by Eshel and Feldman⁴⁰ has the young altering the relationship between its fitness and the amount of resource it receives, such that the parent's optimum response is to provide more resources leading to an overall increase in offspring fitness. This outcome requires a particular relationship between offspring behaviour and the value of obtaining increased resources (see Box 1). Thus costly begging might be favoured by natural selection if this behaviour led to a marked increase in the value

of extra resources to the young and a consequent increase in parental feeding. The degree to which this mechanism operates in nature will depend on the extent to which the offspring can control the marginal value of parental care.

The second model that involves parental monitoring builds on recent advances in the theory of biological signalling and sexual selection⁴¹⁻⁴³. The offspring is assumed to engage in a behaviour such as begging, the intensity of which uniquely reflects its condition (a state variable that determines the benefits it obtains from extra resources). By monitoring levels of begging, the parent obtains an accurate estimate of offspring condition and uses this information to dispense the parental optimum resource share. Clearly this type of signalling is advantageous for the parent; the problem is to explain why selection acting on the offspring does not lead to cheating and misrepresentation. It can be shown that the signalling system is evolutionarily stable as long as the signal is costly to produce, and as long as the value of obtaining more resources increases with diminishing returns^{44,45}. For young in any particular condition, the benefits of signalling at a greater intensity and gaining more resources

are exactly balanced by a combination of two costs: the direct cost of the signal, and the indirect costs of taking food from relatives. Signalling models predict that the parent obtains accurate information and so can enforce its own resource optima: the parent wins. Yet the necessity that an evolutionarily stable signalling system uses costly signals is a reflection of the underlying parent-offspring conflict and leads to a reduction in parental fitness. A further prediction is that the extent of costly solicitation depends on the variance in that component of offspring condition that cannot be directly assessed by the parent⁴⁴. Uncertainty is likely to be greatest around the time of weaning, which may account for the increase in costly offspring behaviour that Trivers³ termed weaning conflict.

Is there no place for cheating and psychological manipulation in resolution models that involve offspring signalling and parental response? Theoretical studies of signals of quality suggest that some misrepresentation can occur, as long as most of the information obtained by the receiver is accurate⁴⁶. Although not yet formally modelled, similar phenomena are undoubtedly possible when an offspring signals to its parent. For example,

BOX 1 Resolution models of parent-offspring conflict

HERE I describe four resolution models of parent-offspring conflict within a single framework. Consider a parent that rears a single offspring each year. The parent allocates an amount of resources y to the offspring whose fitness is $f(y; \alpha)$, where α is a vector of other parameters that might represent offspring behaviour or condition. The future reproductive success of the parent is $g(y; \beta)$, which depends on y as if assume there is a cost of reproduction. Other factors that may influence future parental fitness are included in the parameter vector β . As the parent is identically related to her current and future offspring, its fitness is $W_p \propto f(y; \alpha) + g(y; \beta)$. Because the offspring is more related to itself than its siblings, its fitness is $W_o \propto f(y; \alpha) + rg(y; \beta)$, where r is the coefficient of relatedness. I assume that $f(y; \alpha)$ increases with y , but with diminishing returns, whereas $g(y; \beta)$ decreases with y , either linearly or with decreasing slope.

Trivers battleground. If the parent has full control of resource division, and if there are no fitness consequences of changing resource share via other parameters ($\partial \alpha / \partial y = \partial \beta / \partial y = 0$), then parental fitness is maximized when $B/C = 1$, where $B = f_y(y; \alpha)$ is the marginal benefit to the current offspring of receiving more food, and $C = -g_y(y; \beta)$ is the marginal cost in terms of lost future reproductive success. If the young are in control of resource distribution then offspring fitness is maximized when $B/C = r$. Comparison of the two optima shows there is a genetic conflict of interests (i.e. a battleground) when $1 < B/C < r$. This is essentially Trivers' result, reformulated to describe the optimum distribution of a continuous resource. A simple graphical illustration of this result is shown in Fig. 1A.

Parker and Macnair resolution. Young engage in a behaviour, x , perhaps the level of begging, which influences the amount of food provided by the parent, and whose level is influenced by the resources received. The expressions for current and future reproductive success are thus $f(y, x)$ and $g(y)$, respectively. The resolution of the conflict is specified by an ESS pair $\{y, x\}$ that is immune to invasion by alternatives y' or x' . The ESS is found by differentiating fitness with respect to behaviour, setting $\partial W_p(y') / \partial y'_{y=y} = 0$ and $\partial W_o(x') / \partial x'_{x=x} = 0$, and solving for y and x . Parker and Macnair assumed that near the ESS the marginal increase in resources due to higher levels of solicitation is $dy/dx' = y/x$. The form of this function implies that a fixed increase in solicitation would (1) lead to a relative rather than an absolute increase in resources (hence dy/dx' is a function of y), and (2) be less efficient when solicitation levels are already high (hence dy/dx' is an inverse function of x). The marginal reduction in solicitation consequent on increased resource supply is $dx/dy' = x/y$. This function embodies the assumptions that a fixed increase in resources (1) leads to a relative rather than an absolute drop in the level of solicitation, and (2) has less effect when resource levels are high. At the ESS, the resolution share of resources is given by the expression $B/C = 0.5(1+r)$, which is intermediate between the parent and offspring optima (Fig. 1B).

Yamamura and Higashi resolution. Here the resolution is determined by costly parental (z) and offspring (x) behaviours. The expressions for

current and future reproductive success are thus $f(y, x)$ and $g(y, z)$, respectively. At the ESS, it is assumed that a deviation in behaviour by one party, say the offspring ($x \rightarrow x + \delta$), which enables it to get more food ($y \rightarrow y + \Delta$) leads to an immediate deviation by the parent ($z \rightarrow z + \delta$), the latter being the minimum response required to counter the demand for more food. At the ESS the marginal costs and benefits of the behavioural deviation exactly cancel each other out: $\Delta[f_y(y, x) + g_y(y, z)] + \delta[f_x(y, x) + rg_z(y, z)] = 0$. It is assumed that the ratio of the costs $g_z(y, z)/f_x(y, x)$ is a constant k . When $k < 1$ the parent finds it relatively cheap to manipulate offspring behaviour, and when $k > 1$ the reverse is true. A similar expression can be derived for deviations in parental behaviour and Δ/δ eliminated to obtain the ESS condition $B/C = (2rk + 1 + r)/[k(r + 1) + 2]$, which again can be compared with the offspring and parental optima (see above). Under this model, the optimum resource division is always intermediate between the offspring and parental optima, being nearest (but not equal to) the offspring optimum when $k = \infty$, and nearest (but not equal to) the parental optimum when $k = 0$ (Fig. 1B).

Eshel and Feldman resolution. For this mechanism, we need to consider one extra parameter, x , again representing offspring behaviour. Whereas offspring behaviour influences the parent, the reverse is not true. The ESS resolution is given implicitly by two equations: $f_y(y, x) + g_y(y) = 0$, and $f_x(y, x) + (dy/dx)[f_y(y, x) + rg_y(y)] = 0$. Eshel and Feldman assume that a costly offspring behaviour ($f_x(y, x) < 0$) changes the parental optimum ($f_{yx}(y, x) \neq 0$) so that the young gains an overall increase in fitness. The ESS is found by substituting $dy/dx = -f_{yx}(y, x)/[f_{yy}(y, x) + g_{yy}(y)]$ (the parental optimum response) and solving the two conditions simultaneously. The offspring changes the marginal benefits of resource provision from B to B' and from C to C' because the offspring has higher fitness when $B'/C' = 1$ than when $B/C = 1$ (Fig. 1C).

Signalling resolution. We now need to consider two extra parameters, offspring behaviour (x) and offspring condition (c) so that $f(y; \alpha) = f(y, x, c)$. Offspring in good condition obtain less benefit from extra resources ($f_{yc}(y, x, c) < 0$). Assume that the parent can assess condition by monitoring offspring behaviour (the signal). The parental optimum, which depends on c , is given implicitly by $f_y(y, x, c) + g_y(y) = 0$, an expression of the form $B/C = 1$. The behaviour of the offspring will be evolutionarily stable if $f_x(y, x, c) + (dy/dx)[f_y(y, x, c) + rg_y(y)] = 0$. The first term represents the marginal costs of signalling for an offspring in condition c , and the second term the consequences of resource redistribution that depend on dy/dx , the response of the parent to increased signalling. The ESS resolution is found by combining the two expressions and solving the resultant differential equation with an initial value provided by the observation that, if signalling is honest, young in top condition gain nothing by signalling. Note that if signalling influences the marginal value of extra resources ($f_{yx}(y, x, c) \neq 0$), then the resolution also involves an element of the Eshel & Feldman mechanism.

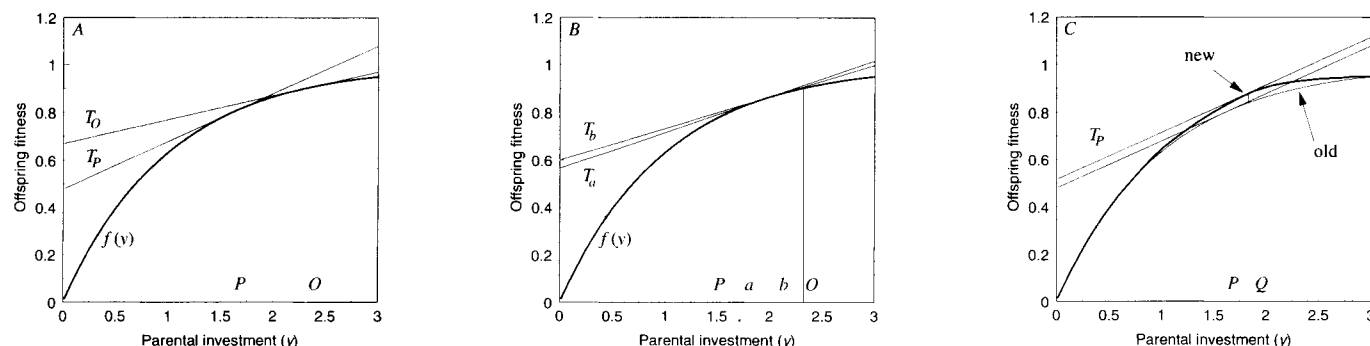


FIG. 1 A graphical representation of the battleground and resolution of parent-offspring conflict. A, The battleground. I assume that offspring fitness is a function $f(y)$ of the resources it receives (y). The parent suffers a linear reduction in future reproductive success as it invests more in its current offspring (that is, future reproductive success is $g(y) = G + \gamma y$, where G is a constant and γ describes the marginal cost of parental investment). The assumption of a linear cost of reproduction is undoubtedly artificial although heuristically valuable in allowing a simple graphical solution. The optimum level of investment from the parent's point of view occurs when the marginal benefits of feeding the current young equal the marginal costs in terms of lost future reproductive success (when $df(y)/dy = \gamma$): this is the amount of investment (P) at which a line with slope $\gamma(T_P)$ in the figure) is just tangent to the offspring fitness curve ($f(y)$). From the point of view of the offspring, the marginal cost of parental investment is $r\gamma$ and its optimum level of investment (O) occurs when a line with slope $r\gamma(T_O)$ is just tangent to the offspring fitness curve. The magnitude of the difference between P and O is a measure of the extent of potential parent-offspring conflict. B, The resolution of the conflict in the Parker-Macnair model occurs when the marginal benefits of feeding the current young equal

$0.5(1+r)\gamma$ (see Box 1): this is the point (a) when a line with that slope (T_a) is just tangent to the offspring fitness curve. The Yamamura-Higashi model specifies a range of possible outcomes (determined by their parameter k ; see Box 1): if the parent finds it relatively cheap to manipulate offspring behaviour then the resolution occurs at the same point as in the Parker-Macnair model; if the reverse is true then the resolution occurs at point b , nearer the offspring optimum. C, In the Eshel-Feldman model, the offspring changes the shape of the curve $f(y)$ relating fitness to parental investment such that the parental optimum investment ($Q > P$) is now higher (the tangent of the 'new' curve with a line of slope $\gamma(T_P)$ occurs at a higher level of investment than the tangent of the same slope with the 'old' curve). Although the offspring benefits from increased investment, overall the parent suffers a reduction in inclusive fitness, as in achieving its new optimum it has to reduce its future reproductive success. In the signalling model, as long as the costs of signalling are uncorrelated with investment, the resource received by the offspring is at the parental optimum. To draw the graphs I set $f(y) = 1 - \exp(-cy)$ with $c=1$, $\gamma=0.2$ and $r=0.5$. The 'new' curve in C was drawn by hand.

if offspring differ in the costs they experience when soliciting resources, those that find signalling relatively cheap or expensive will signal at above or below average levels respectively, leading to the parent providing them with slightly more or less resources than is appropriate for their condition. The important point is that cheating can only occur if the signalling system is largely accurate, and cheating is not required to explain costly solicitation.

Dynamics of conflict

An important simplification of all the resolution models discussed so far is the treatment of parental care as a series of events that have independent consequences for offspring fitness. In fact, the amount of resources provided for the young today will influence its behaviour tomorrow, as well as the effort the parent will have to expend in finding resources on the following day. Embedding the models discussed in Box 1 within a dynamic optimization framework is the obvious way to proceed. Unfortunately, dynamic games of this complexity are notoriously hard to study, analytically or numerically⁴⁷, and no models of this type have been attempted so far. However, two studies have looked at a simpler dynamic problem inspired by sea-bird biology^{48,49}. Each day the offspring has the choice of either fledging or remaining in the nest, and the parent must choose whether or not to feed the chick. The different decisions are influenced by a variety of factors, many of which have been measured in the field, and the resolution of the game is determined by dynamic programming⁴⁷. One unusual aspect of the biology of some sea birds, that nestling weight peaks some days before fledging, emerged as a prediction of the model. A criticism of this work is the use of a very simple expression for offspring fitness that excludes the contribution of future siblings to its inclusive fitness.

A different aspect of the dynamics of parent-offspring conflict is the use of punishment to mould offspring behaviour^{1,3}, a subject recently discussed by Clutton-Brock and Parker⁵⁰ in the context of a broader survey of punishment in animals. Suppose the young engages in behaviour that increases its own fitness at the expense of a parent or sibling. If this is a one-off event, the parent will be selected to respond in a way that maximizes its fitness given the young's behaviour. However, if the opportunity for the young to be selfish recurs then the parent may be selected to penalize the offspring by a fitness-reducing punishment. Although punishing the young also reduces parental fitness, in the long term the parent will gain if punishment deters the young from repeating the action. Parental punishment will succeed as long as the young has the ability to change its behaviour in the light of parental retaliation, an ability we might reasonably expect to evolve. Although formal models have yet to be explored, incorporation of punishment may limit the applicability of the Parker-Macnair and Eshel-Feldman mechanisms described in Box 1, as both resolution models assume that the parent reduces its fitness by making a one-off response to the behaviour of its offspring. It is also possible that young may attempt to 'counter punish' their parents, leading to interactions that resemble the Yamamura-Higashi model. There is an obvious need for more study of this problem.

Large families

Most birds and mammals do not produce offspring singly but in litters or clutches of two to many young. Family life is now complicated by sibling conflict as well as parent-offspring conflict³¹. In animals without parental care, interactions between siblings are the sole determinants of the resolution of sibling conflict. These interactions frequently take the form of exploitation or interference competition and are fairly straightforward

to model. A common prediction is that sibling competition leads to inefficient resource utilization or wasteful conflict, as young attempt to pre-empt resources^{26,27,34,52}. However, unlike the so-called tragedy of the commons, which occurs when unrelated individuals compete for the same resources, sibling relatedness places a bound on the wastefulness of the competition.

In animals with parental care, sibling conflict can still be resolved by interactions among the young, or its outcome may be determined solely or partly by the parent. In some bird species, young jostle amongst themselves to be nearest the nest entrance where they are fed by the parent^{53,54}. Here the parent seems to allow the young to decide feeding priorities. Many species of birds lay eggs that hatch at different times, leading to marked size hierarchies among the young which influence their relative competitive abilities⁵⁵. In the most extreme case, each offspring may be able totally to dominate its smaller siblings. The largest will thus take its optimum share of the total resources provided by the parent, the second largest its optimum share of what is left, and so on down the hierarchy⁵⁶. Such a model predicts the observed distribution of resources in some species of ardeids (herons, egrets and allies)^{57,58}. The problem with both of these observations is to explain why the parent lets the young decide the feeding priority; presumably any benefits of imposing its own priorities fail to outweigh costs such as the wastage of time.

More frequently, sibling conflict is resolved through the actions of the parent. Most theoretical studies of this problem are battleground models that assume a certain parental response to offspring solicitation, and go on to calculate the optimum behaviour for the young. Several models have assumed that all young have equal competitive ability, and have sought to relate the extent of possible parent-offspring conflict to clutch size^{26,34,51,59,60}. The extent of potential conflict is strongly influenced by whether the costs and benefits of increased solicitation are experienced by the individual or spread throughout the brood. However, whether a battleground model can provide robust, testable predictions is questionable, especially in this case where the relationship between potential conflict and brood size is highly model specific.

Two of the mechanisms described in Box 1 have been extended to the simultaneous resolution of sibling and parent-offspring conflict. Parker and Macnair³⁵ assumed that parents have a fixed amount of resource to distribute among the members of a brood, and provide more food to individuals begging at relatively high intensities. As in the single-offspring model, the resolution is a compromise between the optimum resource allocation for offspring and parent. Signalling models have been analysed for the case of two chicks in a nest⁴⁵. These models predict that begging levels of particular offspring should be influenced not only by their individual condition but also by the condition of their nest mate. Thus if one chick is fed or deprived of food, and in consequence decreases or increases its rate of begging, its nest mate should also decrease or increase its begging rate, although to a lesser extent. There is some experimental evidence in support of this from studies of American robins⁶¹.

Two other types of conflict within the family may also need to be considered. First, where both parents provide resources for the young but do not pair for life, the ESS resources provided by one parent will depend on how hard the other works. Assessing the potential battleground for parent-offspring conflict is now quite complex and needs to take into account the effect of current reproductive effort on the future reproductive success of both parents, when rearing young together and with other partners²³. The resolution of parent-offspring conflict through Parker and Macnair's mechanism has been studied in this context²³. Second, the relatedness of an offspring to its siblings is normally defined as the average of the probabilities that a paternally and maternally derived gene in the offspring is present identically by descent in the sibling. If the sibling shares a mother and not a father then the relatedness of paternally derived genes

is zero. Parent-offspring conflict may thus be complicated by intragenomic conflict. In a series of remarkable studies, Haig and colleagues^{18,19,20} have shown that such considerations can make sense of some curious biochemical and physiological interactions within the womb. The possibility that intragenomic conflict influences behavioural interactions after birth is probably less likely, but should certainly be considered.

Brood reduction

Analyses of the distribution of food and other resources, and of the length of the period of parental care, can make use of the theoretical machinery of kin selection, phenotypic modelling and calculus because the assumptions of weak selection and additive fitness effects can be defended. There are, however, interactions within the family that are decidedly not weak or additive, the most spectacular of which is probably brood reduction, the killing of an offspring by a parent or sibling. Brood reduction is widespread but sporadic throughout the animal kingdom⁵¹ and has been much studied in birds⁶²⁻⁶⁴. In groups such as raptors, boobies, penguins and herons, the youngest chick is often killed, invariably by an elder sibling. Several adaptive explanations for brood reduction have been proposed, including that it allows the parents to capitalize on unpredictable years of resource plenty when the full brood is produced⁶⁵⁻⁶⁸, and that producing extra young is an insurance against the death or infertility of eggs⁶⁹⁻⁷¹. Among invertebrates, within-brood cannibalism is common and siblicide often may provide direct trophic benefits^{4,72}.

To what extent are there conflicts of interest between parent, surviving offspring and victim? Simple considerations of costs, benefits and relatedness suggest that survivors, parents and victims should be progressively less willing to countenance brood reduction but, as O'Connor⁶² has stressed, even the victim may be selected to destroy itself if the benefits to its relatives are high enough. Battleground models of brood reduction confirm this suggestion^{28,62,73-76} and also highlight the importance of whether the victim is chosen at random from the brood or is an identifiable runt. If the former is the case then higher levels of siblicide may be observed as any gene for non-siblicidal behaviour will be severely penalized when rare because its carrier will tend to be eliminated by brood mates^{8,73,76}. This is not the case when the victim is a runt chosen independently of genotype. Genetic models also predict that, when the victim is chosen at random, alternative stable states are possible and the potential for parent-offspring conflict depends on the history of the population. A more technical point is that the presence of a fixed victim often, although not always, allows the use of phenotypic as opposed to genetic models.

Although conflict is possible, is it actually present in the field and how is it resolved? Where brood reduction has evolved as an insurance policy against infertile eggs there may be no conflict if the rearing of supernumerary young is impossible or very expensive^{70,71,77-79}. In groups where this hypothesis is thought to apply, brood reduction normally occurs through siblicide with no parental interference, which is consistent with the hypothesis of no conflict^{57,63,80}. However, if there were conflict, and if the young always got their way, then the parent might have been selected to allow brood reduction to happen as quickly as possible. Alternatively, allowing young to fight may ensure that the strongest survives^{81,82}. Where parents can rear the complete clutch of eggs there would seem to be more scope for conflict. Forbes⁷⁹ has studied a deterministic resolution model of brood reduction in which the parent responds to a drop in offspring number by reducing the amount of food to be shared by the remaining nest mates (as occurs, for example, in egrets⁸³). The parental response reduces the area of parameter space where parent-offspring conflict occurs, and also reduces the maximum fitness reduction experienced by the losing party. Extensions of these models to include the seasonal dynamics of provisioning,

and brood reduction in a temporally varying environment, are the next steps in producing a testable theory.

Conclusion

The study of parent-offspring conflict is important for several reasons. First, interactions between parents and young are some of the most widespread and basic social behaviours exhibited by animals (and even by plants). Second, the prediction and demonstration of parent-offspring conflict is extremely significant in demonstrating the basic correctness of the 'genes-eye' paradigm of social evolution: such conflicts are utterly inexplicable by alternative approaches. Third, there is obviously enormous interest in human parent-offspring interactions. Behavioural interactions between human parents and their children are likely to be immeasurably more complicated than their animal counterparts, yet an understanding of the underlying evolutionary tensions must be valuable in interpreting human behaviour. Although I have concentrated in this review on behavioural interactions between parents and offspring, the recent demonstration of the role of genetic conflicts of interest in determining the biochemical interactions between the fetus and mother¹⁸⁻²⁰

has direct applications to human biology and even human medicine.

Nevertheless, the importance of parent-offspring conflict in moulding animal behaviour is still fiercely debated. In the past, much of the theory in the field has been devoted to the demonstration of potential conflicts: defining the battleground or, to quote Mock and Forbes⁵, the performance of 'elaborate logic checks'. Because these models are not designed to predict what should be observed in nature, they are very hard to test. The last few years have seen a switch in emphasis to models that attempt to predict the resolution of parent-offspring conflict, as well as a series of new ideas about how the resolution may be brought about. Although major theoretical lacunae still exist, in particular the lack of a dynamic component to parental care, there is now a real prospect of a conceptual framework that will generate detailed hypotheses that can be tested in the laboratory and in the field. □

H. C. J. Godfray is in the Department of Biology and NERC Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

- Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton Univ. Press, Princeton, NJ, 1991).
- Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
- Trivers, R. L. *Am. Zool.* **14**, 249-264 (1974).
- Alexander, R. D. *A. Rev. Ecol. Syst.* **5**, 325-383 (1974).
- Mock, D. W. & Forbes, L. S. *Trends Ecol. Evol.* **7**, 409-413 (1992).
- Bateson, P. *Trends Ecol. Evol.* **9**, 399-403 (1994).
- Trivers, R. L. & Hare, H. *Science* **191**, 249-263 (1976).
- Godfray, H. C. J. *Am. Nat.* **129**, 221-233 (1987).
- Nonacs, P. *Q. Rev. Biol.* **61**, 1-21 (1986).
- Boomsma, J. J. *Am. Nat.* **133**, 517-532 (1989).
- Bourke, A. F. G. *Q. Rev. Biol.* **63**, 291-311 (1988).
- Rosenheim, J. A. *Am. Nat.* **141**, 90-104 (1993).
- Eshel, I. & Sansone, E. *Am. Nat.* **138**, 954-972 (1991).
- Eshel, I. & Sansone, E. *Am. Nat.* **143**, 987-1006 (1994).
- Uma Shaanker, R., Ganeshaiah, K. N. & Bawa, K. S. *A. Rev. Ecol. Syst.* **19**, 177-205 (1988).
- Queller, D. C. *Oxford Surv. evol. Biol.* **6**, 73-109 (1989).
- Emlen, S. T. & Wrege, P. H. *Nature* **356**, 331-333 (1992).
- Haig, D. & Graham, C. *Cell* **64**, 1045-1046 (1995).
- Haig, D. *Q. Rev. Biol.* **68**, 495-532 (1993).
- Haig, D. *Dev. Biol.* **3**, 153-160 (1992).
- Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).
- Johnstone, R. A. *Nature* **372**, 172-175 (1994).
- Parker, G. A. *Anim. Behav.* **33**, 519-533 (1985).
- Lundberg, S. & Smith, H. J. *theor. Biol.* (1994).
- Godfray, H. C. J. & Parker, G. A. *Phil. Trans. R. Soc. B* **332**, 67-79 (1991).
- Godfray, H. C. J. & Parker, G. A. *Anim. Behav.* **43**, 473-490 (1992).
- Roitberg, B. D. & Mangel, M. *Am. Nat.* **142**, 443-456 (1993).
- Parker, G. A. & Mock, D. W. *Evol. Ecol.* **1**, 161-174 (1987).
- Redondo, T. & Castro, F. *Ibis* **134**, 180-187 (1992).
- Haskell, D. *Proc. R. Soc. B* **257**, 161-164 (1994).
- Zahavi, A. in *Evolutionary Ecology* (eds Stonehouse, B. & Perrins, C.) 253-259 (University Park Press, Baltimore, MD, 1977).
- Parker, G. A. & Macnair, M. R. *Anim. Behav.* **26**, 97-110 (1978).
- Macnair, M. R. & Parker, G. A. *Anim. Behav.* **26**, 111-122 (1978).
- Macnair, M. R. & Parker, G. A. *Anim. Behav.* **27**, 1202-1209 (1979).
- Parker, G. A. & Macnair, M. R. *Anim. Behav.* **27**, 1210-1235 (1979).
- Stamps, J., Metcalf, R. & Krishnan, V. *Behav. Ecol. Sociobiol.* **3**, 367-392 (1978).
- Stamps, J. & Metcalf, R. A. in *Sociobiology: Beyond Nature/Nurture* (eds Barlow, G. W. & Silverberg, J.) 589-618 (Westview, 1980).
- Feldman, M. W. & Eshel, I. *Am. Nat.* **119**, 285-292 (1982).
- Yamamura, N. & Higashi, M. *Evolution* **46**, 1236-1239 (1992).
- Eshel, I. & Feldman, M. W. *Am. Nat.* **137**, 167-185 (1991).
- Zahavi, A. *J. theor. Biol.* **67**, 603-605 (1975).
- Grafen, A. *J. theor. Biol.* **144**, 473-516 (1990).
- Grafen, A. *J. theor. Biol.* **144**, 517-546 (1990).
- Godfray, H. C. J. *Nature* **352**, 328-330 (1991).
- Godfray, H. C. J. *Am. Nat.* (in the press).
- Johnstone, R. A. & Grafen, A. *Anim. Behav.* **46**, 759-764 (1993).
- Mangel, M. & Clark, C. W. *Dynamic Modeling in Behavioral Ecology* (Princeton Univ. Press, Princeton, NJ, 1988).
- Clark, C. W. & Ydenberg, R. C. *Evol. Ecol.* **4**, 312-325 (1990).
- Tokuda, H. & Seno, H. *J. theor. Biol.* **170**, 145-157 (1994).
- Clutton-Brock, T. H. & Parker, G. A. *Nature* **373**, 209-216 (1995).
- Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, Oxford, in the press).
- Sjerps, M. & Haccou, P. *Evol. Ecol.* **8**, 269-287 (1994).
- Rydén, O. & Bengtsson, H. Z. *Tierpsychol.* **53**, 209-224 (1980).
- McRae, S. B., Weatherhead, P. J. & Montgomerie, R. *Behav. Ecol. Sociobiol.* **33**, 102-106 (1993).
- Magrath, R. D. *Biol. Rev.* **65**, 587-622 (1990).
- Parker, G. A., Mock, D. W. & Lamey, T. C. *Am. Nat.* **133**, 846-868 (1989).
- Mock, D. W. *Behav. Ecol. Sociobiol.* **20**, 247-256 (1987).
- Mock, D. W., Lamey, T. C. & Ploger, B. J. *Ecology* **68**, 1760-1772 (1987).
- Harper, A. B. *Am. Nat.* **128**, 99-114 (1986).
- Lazerus, J. & Inglis, B. *Anim. Behav.* **34**, 1791-1804 (1986).
- Smith, H. G. & Montgomerie, R. *Behav. Ecol. Sociobiol.* **29**, 307-312 (1991).
- O'Connor, R. J. *Anim. Behav.* **26**, 79-96 (1978).
- Mock, D. W. in *Infanticide: Comparative and Evolutionary Perspectives* (eds Hausfater, G. & Hrdy, S. B.) 3-30 (Aldine, New York, 1984).
- Mock, D. W., Drummond, H. & Stinson, C. H. *Am. Scient.* **78**, 438-449 (1990).
- Lack, D. *Ibis* **89**, 302-352 (1947).
- Lack, D. *The Natural Regulation of Animal Numbers* (Clarendon, Oxford, 1954).
- Temme, D. H. & Charnov, E. L. *J. theor. Biol.* **126**, 137-147 (1987).
- Pijanowski, B. C. *Am. Nat.* **139**, 1270-1292 (1992).
- Dorward, D. F. *Ibis* **103**, 174-200 (1962).
- Anderson, D. J. *Am. Nat.* **135**, 334-350 (1990).
- Forbes, L. S. *J. theor. Biol.* **147**, 345-359 (1990).
- Polis, G. A. *Rev. Ecol. Syst.* **12**, 225-251 (1981).
- Stinson, C. H. *Evolution* **33**, 1219-1225 (1979).
- Dickens, D. W. & Clark, R. A. *J. theor. Biol.* **125**, 301-305 (1987).
- Mock, D. W. & Parker, G. A. *Evolution* **40**, 459-470 (1986).
- Godfray, H. C. J. & Harper, A. B. *J. theor. Biol.* **145**, 163-175 (1990).
- Anderson, D. J. *Evolution* **44**, 2069-2082 (1990).
- Forbes, L. S. *Oikos* **62**, 325-332 (1991).
- Forbes, L. S. *Am. Nat.* **142**, 82-117 (1993).
- Drummond, H., Gonzalez, E. & Osorno, J. L. *Behav. Ecol. Sociobiol.* **19**, 365-373 (1986).
- Simmons, R. *Ibis* **130**, 339-357 (1988).
- Koslowski, J. & Stearns, S. C. *Evolution* **43**, 1369-1377 (1989).
- Mock, D. W. & Lamey, T. C. *Am. Nat.* **138**, 1015-1026 (1991).

ACKNOWLEDGEMENTS. I thank A. Burt, T. Clutton-Brock, A. Grafen, D. Haig, A. Kacelnik, D. Mock, G. Parker and M. Rees for discussions.

Propagation of solar oscillations through the interplanetary medium

David J. Thomson, Carol G. MacLennan & Louis J. Lanzerotti

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Time-series analysis of the fluxes of interplanetary charged particles measured by the Ulysses and Voyager spacecraft reveals many periodic components. From 1 to 140 μHz , the spectral components are consistent with those estimated (but not confirmed) for gravity-mode oscillations of the Sun: from 1,000 to 4,000 μHz , the spectral lines closely match the frequencies of known solar pressure modes. These concordances imply that the solar wind and the interplanetary magnetic field transmit solar oscillations and thus might be used to probe the interior structure of the Sun.

THE Sun oscillates, in much the same way as a drum or a ringing bell, with various modes of oscillation. The pressure or p modes are acoustic standing waves, which have been thoroughly mapped using optical techniques. The gravity or g modes, on the other hand, result from gravitational restoration of density gradients within the Sun. The g modes should be sensitive indicators of the structure deep within the Sun, particularly in the core, about which we have little direct information. The detection of g -mode oscillations has been reported¹, but not been confirmed (P. H. Scherrer, personal communication). Between the lack of observations and the modes' sensitivity to density gradients, the possible frequencies of g modes² are much less certain than those of p modes.

Although g -mode oscillation frequencies are not currently included as constraints in models of the solar interior³, they may play an important role in mixing the hydrogen fuel and helium fusion products at the centre of the Sun, and thus may be relevant to the problem of the missing solar neutrinos²⁻⁴. Models of the solar interior predict more neutrinos than are observed^{4,5}. This long-standing discrepancy strikes at the heart of our understanding of stellar structure and basic nuclear physics.

We have analysed the time series of fluxes of low-energy interplanetary charged particles, whose ultimate origin is the Sun. Based on time-series analysis of data from the Voyager II spacecraft (during 1985) and from the Ulysses spacecraft (1992-94), we find evidence that solar oscillations can be observed in the variations of the particle flux. We were surprised to find evidence for periodic components other than solar rotation in these data. Traditionally, time variations in the particle fluxes have been ascribed solely to sources such as the injection and propagation of particles from solar disturbances (for example, flares), the release of particles from planetary magnetospheres, the acceleration of particles at interplanetary shock waves and fluctuations in the magnetic field intensity and direction in interplanetary space. All of these effects^{6,7} imply a continuous spectrum, not discrete features.

The Sun's control over the interplanetary environment, through its magnetic field and the solar wind, makes the Sun a likely source of the observed periodicities. The prime candidates in our observed frequency range of 1-140 μHz are g modes that are transmitted into the interplanetary medium. We found that the observed oscillation in the particle flux had the characteristic stability and nearly equally spaced periods expected for g modes⁸. We also looked for, and found, evidence for solar p -mode oscillations in the 1,000-5,000 μHz range.

What physical mechanism could explain this apparent modulation of the solar wind by solar oscillations? We suggest that supergranulation motions, which have been assumed to be

random motions on the solar surface, are not completely random, but partially the effect of many g modes. It has been predicted⁹ that magnetic flux 'frozen' in the supergranulation is responsible for the observed transverse magnetic fields in the polar regions of the Sun¹⁰. Periodic components in the interplanetary magnetic field would cause periodic modulation of the charged-particle flux.

Solar modes

Viewed from a sidereal reference frame, the surface motion of a p - or g -mode oscillation is given approximately⁸ by

$$a_{nl}P_l^m(\cos\theta)\cos(m\phi-2\pi ft) \quad (1)$$

where f is the frequency of oscillation, a_{nl} is the amplitude depending on the spherical harmonic degree l and radial order n , and P_l^m is the associated Legendre function of order m . The solar longitude is given by ϕ , and θ is the colatitude. The frequencies of the p modes have been mapped thoroughly using optical Doppler methods^{11,12} for frequencies in the range 1,000-5,000 μHz (periods 4-20 minutes). The ratio of circulating energy to energy dissipated per cycle (Q) is about 10,000, so the amplitude of the p modes stays fairly constant over several days¹³. Our understanding of the convection zone of the Sun results largely from models that reproduce the observed p modes¹⁴.

The g modes have frequencies less than about 400 μHz (periods >40 minutes). They are expected¹⁵ to have extremely high values of Q , so their amplitudes should stay relatively constant for thousands of years.

Data preparation and analysis

Table 1 lists the different sets of interplanetary data^{16,17} that we analysed, giving the particle species, channels and time-intervals. We selected the data sets by several criteria: data continuity, location of the spacecraft within or outside the heliospheric current sheet, and the presence or absence of distinct solar-particle events. For the most part, data were available as hourly averages, resulting in a minimum observable period of 2 hours, or a Nyquist frequency, $1/2\Delta t$, of 139 μHz . (Because averaging is not a good filter, some frequencies close to the band edge may be aliases.) Many of the processes involved in accelerating the solar wind are multiplicative¹⁸, so, for most of the analyses, logarithms of the hourly flux values were used.

The more lengthy data sets all have short gaps due to tracking omissions, instrument calibrations, telemetry noise and so on. We used a two-pass procedure (formally an EM algorithm¹⁹) that has been proven analytically to be statistically consistent in missing-data problems. This spectrum estimation method corres-

TABLE 1 Data sets used in analysis

Group	Number of freq.	Source				
		Rate	Spacecraft*	Channel	Species	Dates #
1	139	Hourly	Ulysses†	P2' 80–130 keV per nucl.	Ions	220–290 (1992)
2	129	Hourly	Ulysses†	W1 480–996 keV	H	220–290 (1992)
3	126	Hourly	Ulysses†	W3 389–1,278 keV per nucl.	He	220–290 (1992)
4	117	Hourly	Ulysses‡	W1 480–996 keV	H	42–256 (1993)
5	129	Hourly	Ulysses‡	W2 968–1,204 keV	H	42–256 (1993)
6	144	Hourly	Ulysses‡	W3 389–1,278 keV per nucl.	He	42–256 (1993)
7	38	Hourly	Voyager 2§	Ch 1 0.52–1.45 MeV	H	79 (1985)–19 (1986)
8	200	Hourly	Ulysses	DE1 38–53 keV	e	1–245 (1994)
9	200	Hourly	Ulysses	W1 480–996 keV	H	1–245 (1994)
10		1-min	Ulysses¶	DE1–DE4	e	228–243 (1992)
11		1-min	Ulysses¶	E'1–E'4	e ⁻	228–243 (1992)
12	200	Hourly	Imp8, ISEE3	B _x , B _y , B _z , B	—	225 (1978)–32 (1980)
13	200	Hourly	Imp8, ISEE3	B _x , B _y , B _z , B	—	32 (1980)–315 (1981)

B_x, B_y and B_z are the components of the interplanetary magnetic field in geocentric–solar–ecliptic (GSE) coordinates; |B| is the total field intensity. IMP8 and ISEE3 are interplanetary spacecraft; the orbit of IMP8 is Earth-centred and the orbit of ISEE3, at the time of the measurements, was located around the L1 lagrangian point.

* Locations of spacecraft as follows (given as heliographic latitude, distance in AU): † ~15–19 °S, 5.2–5.3; ‡ ~25–40 °S, 4.3–5.0; § ~0 °S, 16.5–19.5; || ~48–80 °S, 2.4–3.8; ¶ ~16 °S, ~5.3.

Dates given as 'day (year)', where day 1 is 1 January.

ponds to the robust procedures developed in refs 20–24, so there is much empirical evidence for its efficacy. Such multi-taper spectrum estimation has been found to be extremely accurate for terrestrial normal-mode seismology^{25–27}.

Our first pass used linear interpolation followed by a five-point running midmean smoother. A spectrum was then com-

puted using a multiple-taper method^{28–30}. Typically a time–bandwidth product of five with eight prolate spheroidal data windows was used for each time series. The harmonic *F*-test^{28,29} was used to identify potential periodic terms in the spectra. The *F*-statistic is the ratio of the energy explained by a single sinusoidal component at a given frequency to the rest of the energy within the selected bandwidth. It is a likelihood-ratio test for the presence of a periodic term against the null hypothesis of a locally white spectrum. The frequency where the *F*-statistic is maximized is, as an estimator of line frequency, close to theoretical limits³¹.

The tentative set of periodic components identified by the *F*-test (typically 100–200), often accounting for more than 80% of the variance, was then used to form a purely periodic reconstruction of the data. This reconstruction was then used as a time-varying mean value function in the data gaps, and the residual spectrum (the non-harmonic portion) was used to generate the interpolators. For the second pass, data gaps of less than 20 samples were filled by a standard Wiener least-squares procedure that used the autocorrelations obtained by Fourier transforming the residual spectrum to interpolate from available good data samples on both sides of the gap. The few longer gaps were filled by linear interpolation.

The spectrum estimation procedure was then repeated on the newly interpolated data. Initial frequency estimates were found with a fast Fourier transform as before, but a search procedure was used to find the frequency giving the maximum *F*-statistic. In addition, this procedure was jackknifed by deleting each window in turn from the analysis³². By this means, non-parametric variances of frequency, phase and amplitude were obtained in addition to the parameter estimates themselves.

In making the above spectral estimates, the detection threshold for periodic elements was set at the 95% level. This level is a compromise between false detections and the detection of actual periodic modes. Because the measurement noise is presumably independent in the different data sets, individual false detections can be suppressed by requiring that the estimated frequencies coincide in the different data sets to assert that a periodic component is present. We rejected frequency estimates with jackknife standard deviations exceeding 0.05 μHz . The discrete frequencies estimated from the different data sets were merged and sorted. Finally, to reduce the number of false detections, we retained only those groups of frequencies that were observed three or more times within a range of 0.2 μHz in the different data sets.

Low-frequency modes

Using Ulysses and Voyager II particle data only, there were 161 such groups of frequency estimates. The 35 groups with four or more coincidences are listed in Table 2.

TABLE 2 Coincident frequencies

N	Groups	Freq. (μHz)	fsd (μHz)	Amp.	F- test	t- test	VR	M1	M2
7	9,3,2,6,5,1,8	0.446	0.004	415.01	17.0	12.6	0.8	1	15
4	6,4,5,8	0.892	0.006	290.51	5.7	4.5	0.8	5	15
4	4,5,6,8	1.797	0.013	92.43	3.8	1.7	0.2	4	14
4	6,4,9,3	2.279	0.013	76.59	4.1	9.2	3.9	1	14
4	4,5,2,3	4.940	0.013	29.51	4.9	6.3	1.8	3	18
4	2,5,3,6	5.254	0.011	30.74	6.5	3.3	0.4	1	11
4	9,8,6,1	6.045	0.004	19.10	11.4	23.8	12.0	6	15
5	5,8,4,6,1	10.349	0.009	16.82	9.2	8.5	1.6	3	15
4	6,3,9,2	10.621	0.016	12.82	5.0	3.5	0.7	2	20
5	8,4,3,1,2	11.114	0.015	12.68	4.6	4.7	0.8	3	17
4	6,5,4,9	12.144	0.005	13.76	7.1	7.7	1.6	3	16
4	8,3,4,9	16.678	0.005	5.48	6.7	18.1	4.4	2	15
4	9,1,3,2	26.201	0.006	7.47	12.5	4.9	1.2	3	11
4	1,5,8,9	27.435	0.004	5.70	6.8	14.3	4.2	2	13
4	3,4,2,6	30.840	0.011	5.21	6.9	4.4	1.0	7	15
4	3,2,5,7	35.456	0.009	4.29	7.5	9.5	5.6	4	22
4	2,8,3,4	46.296	0.005	2.89	9.2	10.9	3.4	3	20
4	9,8,3,5	47.082	0.005	1.57	4.7	12.5	3.0	6	20
4	5,4,2,6	49.577	0.015	2.59	3.5	3.6	0.6	4	19
5	6,4,8,9,5	55.364	0.004	1.80	5.9	11.8	1.0	4	19
4	9,3,6,1	55.901	0.005	2.60	10.6	8.8	2.0	6	19
4	5,2,1,9	56.798	0.005	2.17	7.3	11.1	2.3	2	16
4	9,5,1,8	79.795	0.005	1.43	4.6	16.7	3.5	3	20
4	1,3,4,8	81.038	0.007	1.51	6.0	7.2	1.8	2	14
5	4,6,3,2,1	89.113	0.012	1.66	6.0	5.6	0.8	0	12
4	9,8,1,2	89.578	0.005	1.15	5.1	8.1	1.5	1	8
5	8,7,6,3,9	95.208	0.004	1.22	5.3	20.3	4.6	5	15
4	2,1,5,4	96.524	0.013	1.83	5.2	10.1	3.7	3	16
4	8,9,5,1	97.460	0.009	0.96	4.8	11.8	3.0	3	14
4	9,2,1,8	101.322	0.003	1.09	7.0	14.8	2.2	2	13
4	2,1,9,4	106.768	0.005	1.25	6.6	20.4	3.9	0	9
4	8,9,6,3	111.938	0.004	1.21	5.4	18.2	5.8	4	19
5	7,2,5,3,4	114.042	0.006	1.78	6.5	14.8	4.2	3	17
4	1,8,2,5	124.925	0.006	1.15	7.2	19.7	5.5	2	10
4	2,1,3,8	131.989	0.005	0.78	6.1	9.8	1.9	3	12

N, number of coincident frequencies; Freq., average frequency; fsd, standard deviation of average frequency; Amp., average amplitude, $1,000 \log_{10}(\text{flux})$; *F*-test, geometric mean of the individual detection *F*-statistics; *t*-test, ratio of the range of frequencies to their standard deviation; VR, ratio of the variance computed from individual frequency estimates to the average of the jackknife variances; M1, number of matches with the OMNI magnetic data allowing ± 1 cycle yr^{-1} and ± 2 cycles yr^{-1} and satellite-switching frequency splittings; M2, as M1 plus $\pm 0.558 \mu\text{Hz}$ splittings.

Could the observed frequencies have arisen purely by chance with no deterministic structure present? We take as our null hypothesis, H_0 , the standard assumption that the various time series considered have a purely continuous spectrum with no discrete components apart from solar rotation. The following paragraphs describe four tests of H_0 . These four tests are independent under H_0 , and all four decisively reject the continuous spectrum hypothesis.

First, under H_0 , the number of frequency estimates in fixed frequency intervals should have a Poisson distribution; a χ^2 test for goodness-of-fit of the observed number of intervals containing 0–6 frequency estimates (solar rotation frequencies were excluded) to those expected with random data has $\chi^2 = 98.7$ with 7 degrees of freedom, strongly rejecting the continuous spectrum hypothesis.

Second, the false-detections of the F -test should be uniformly distributed over the Nyquist band. A Kolmogorov–Smirnov³³ D -test for uniformity of the observed frequencies across the Nyquist band (the maximum difference between the empirical and theoretical cumulative distributions) had a value of 0.262. Such a value would occur in a uniform distribution in about 4 out of 10^{74} trials. If the frequencies are uniformly distributed, their spacings should have an exponential distribution³⁴. Figure 1 is a histogram of all frequency differences less than $0.2 \mu\text{Hz}$ from the 1,032 frequency estimates. It is obvious that the distribution is not an exponential, but a multi-modal distribution. Similarly, t -tests (Table 2, column 7), the ratio of the range of frequencies within each cluster to their estimated standard deviation, are unusual, suggesting a bi- or multi-modal distribution.

Third, we compared the values of the F -statistics with those expected from a central F distribution. On the 1,032 frequency estimates, 234 exceeded the 99.5% level, more than twice the 103 expected, a 12σ event. The number above the 99.99% level is 9.9 times expected.

Fourth, under H_0 , the range within each group gives a maximum-likelihood estimate of the width of the search interval, and the sampling properties of such estimates are well known³⁵. The

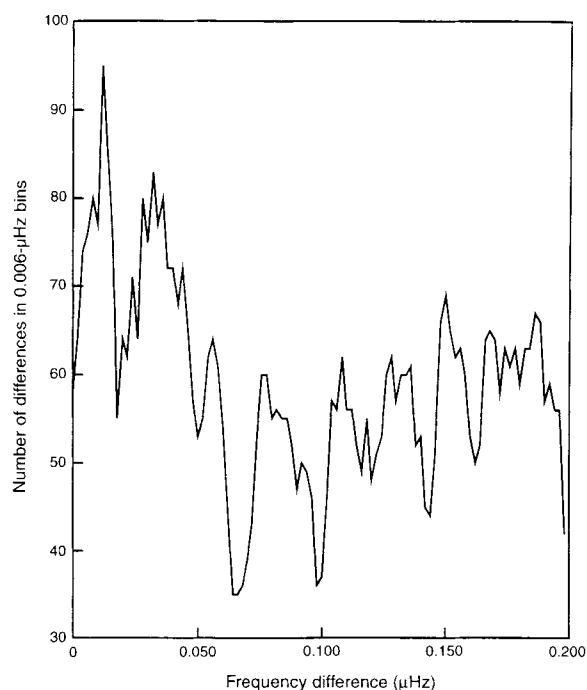


FIG. 1 Histogram (0.006 μHz bins offset 0.003 μHz) of the frequency differences less than $0.2 \mu\text{Hz}$ between all 1,032 raw frequency estimates.

average range with the groups of four and more was $0.074 \mu\text{Hz}$, only 0.37 of the $0.2 \mu\text{Hz}$ search range, again a 14σ rejection of uniformity.

We also examined the time dependence of the amplitude and phase of these apparently periodic components using complex demodulation³⁶. If the observed periodic terms are due to high- Q g modes, then their amplitude and phase should be constant except for a heliographic latitude and longitude dependence from equation (1). Figure 2 shows the outputs of narrow-band filters centred at $f = 105.3732 \mu\text{Hz}$ and $24.1117 \mu\text{Hz}$ as functions of time. Between 1992 and 1994 the phase at $105.3732 \mu\text{Hz}$ (Fig. 2) changed by 2.3×10^6 degrees, while the r.m.s. departure from linear is about 33 degrees. Because there are no zeros in amplitude and no 180-degree phase steps during this time, either a mode with $l=1$ or a sectorial, $l=m$, mode could produce such a result. White noise passed through this filter should have about eight zeros during this time interval, so the observed phase-coherence would be a rare event³⁷ if the source were noise or a low- Q mode. Moreover, where coincident frequencies occur in both the 1985 Voyager data and the 1992 and 1993 Ulysses data,

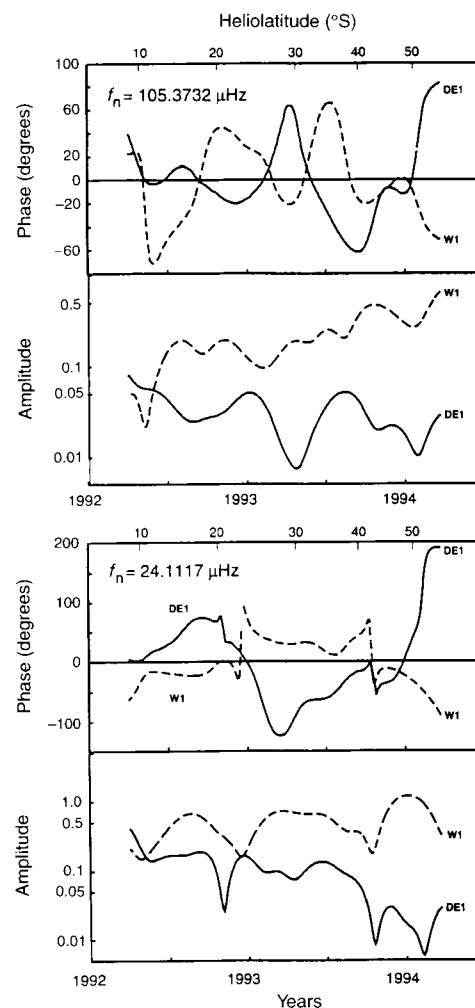


FIG. 2 Phase of complex demodulates on electron channels DE1 and W1 for the interval January 1992 to June 1994 for modes at frequencies of $105.3732 \mu\text{Hz}$ (top) and $24.1117 \mu\text{Hz}$ (bottom). The heliographic latitude of Ulysses is shown at the top of the figure. The latitudes where the coincident amplitude nulls and 180° phase jumps occur imply (equation (1)) that the $24.1117\text{-}\mu\text{Hz}$ line may correspond to an $l=6$ mode. The phase between the discontinuities in late 1992 and late 1993 has been offset by 180° . The absence of these implies that the $105.3732\text{-}\mu\text{Hz}$ line probably is from a mode with $l=1$, $m=0$, so n may be 5 or 6.

the estimated amplitudes have been statistically identical. We did similar tests at all the frequencies listed in Table 2: their amplitude and phase characteristics are consistent with originating in high- Q g modes.

A histogram of differences between the periods corresponding to the 1,032 frequencies are compatible with the differences expected for g modes^{3,8}, but the frequency splitting, see Fig. 1, makes unambiguous identification difficult. The frequency range, period differences and high Q s are the main reasons to suspect that the observed frequencies are the result of g modes.

One frequency that does not appear in our analysis is the '160-minute' oscillation; its frequency was reported³⁸ as 104.1922 μHz sidereal. Our closest group to this frequency, at 104.112 μHz , was not significant enough to be included in Table 2 but even so is 5.3σ away. This is in agreement with the results reported in ref. 39, where the '160-minute' oscillation was not detected in geomagnetic data although it had been claimed^{40,41} to be present.

High-frequency modes

To examine further the hypothesis that solar modes are being observed in the interplanetary medium, we used similar procedures to analyse two eight-day segments, from days 228–243 of 1992, of eight electron channels, E'1 to E'4 and DE1 to DE4, from the Ulysses HI-SCALE instrument¹⁷ at one-minute resolution. Collectively, these channels cover the range 42–315 keV. These were analysed as before, retaining only frequencies where the level of the F -statistic exceeds the 95% point and with jack-knife standard deviations less than 1.6 μHz . For the frequency range 0–5,000 μHz , the 16 data sets (cumulatively about 184,000 samples) gave about 3,300 frequency estimates. As with the low-frequency modes, the hypothesis that the underlying spectrum is continuous is untenable.

Performing similar grouping operations as before, we obtained 602 groups of three or more coincidences. The most striking feature was a strong line at 1.12 ± 0.05 μHz found in 13

of the 16 electron data sets. Our initial suspicion was that this was associated with the spacecraft, but the failure to detect signals of the same frequency in the ion channels argues otherwise. Similar F -tests for periodic components in the daily sunspot number series, the Penticton 10.7-cm solar flux, and the geomagnetic indices K_p and A_p , all show lines significant above the 99% level within the sidereal frequency range 1.1100–1.1115 μHz (periods 10.41–10.43 days). This suggests that the 1.11- μHz oscillation is of solar origin. This frequency is twice that attributed to rotation of the solar core¹³ in optical measurements of p modes. Whatever the source, an amplitude of 0.175 on the $\log_{10}$ electron flux scale results in a peak-to-peak change in the flux of a factor of 3, strong enough to modulate other modes and cause frequency splitting.

We took the 118 optically observed^{11,12} p -mode frequencies between 1,587.5 and 4,199.6 μHz and $0 \leq l \leq 6$ and compared them to our group frequencies. We found that the Ulysses frequencies were higher than the optical frequencies by a factor of about 1.00078. This ratio resulted in matches to 90 groups from the Ulysses data with a tolerance of 1.2 μHz . Fixing the frequency shift at the value determined by the low- l modes, and examining agreements with higher- l modes suggests that some modes with l up to about 25–30 are being seen. Table 3 shows a subset of the results of this matching process arranged by spherical harmonic indices l and n for $0 \leq l \leq 3$ and $15 \leq n \leq 25$ with lists of both optical and Ulysses frequency estimates and standard deviations. Clearly the match is good; the disagreement between the optical and the Ulysses frequencies is often less than the differences between optical determinations.

The frequency shift of 1.00078 is not understood. Ulysses was approaching the Sun at about 1.5 km s^{-1} when these data were taken, so the shift is in the right direction to be a Doppler shift. But if so, the implied speed of 1,900 km s^{-1} is too high for the solar wind (about 600 km s^{-1}), and too low for the nominal electron velocity.

TABLE 3 Comparison of p -mode frequencies (μHz)

Order n		Spherical harmonic degree							
		$l=0$		$l=1$		$l=2$		$l=3$	
15	Opt.	2,228.53	0.09	2,291.97	0.15	2,352.04	0.16	2,407.87	0.28
	Uly.	2,227.40	0.17	2,292.08	0.17	2,352.16	0.17	—	—
16	Opt.	2,362.82	0.05	2,425.53	0.05	2,485.70	0.14	2,541.68	0.17
	Uly.	2,361.83	0.17	2,425.48	0.20	2,486.86	0.14	2,542.74	0.11
17	Opt.	2,496.19	0.09	2,559.22	0.08	2,619.69	0.13	2,676.30	0.15
	Uly.	2,497.45	0.12	2,560.25	0.33	2,621.33	0.21	2,675.66	0.13
18	Opt.	2,629.51	0.09	2,693.41	0.08	2,754.68	0.13	2,811.75	0.05
	Uly.	2,629.95	0.29	2,693.57	0.21	2,755.14	0.14	2,811.18	0.17
19	Opt.	2,764.11	0.08	2,828.20	0.10	2,889.85	0.13	2,947.00	0.12
	Uly.	—	—	2,827.06	0.12	2,890.58	0.13	2,947.03	0.13
20	Opt.	2,899.07	0.08	2,963.35	0.07	3,024.87	0.12	3,082.62	0.12
	Uly.	2,901.28	0.17	2,963.20	0.15	3,023.36	0.20	3,082.50	0.14
21	Opt.	3,033.83	0.06	3,098.28	0.08	3,160.05	0.08	3,217.78	0.24
	Uly.	3,035.52	0.09	3,099.05	0.24	3,160.02	0.21	—	—
22	Opt.	3,168.76	0.05	3,233.26	0.07	3,295.07	0.15	3,354.27	0.32
	Uly.	3,168.84	0.11	3,233.59	0.11	—	—	3,354.50	0.14
23	Opt.	3,303.49	0.09	3,368.69	0.11	3,430.75	0.15	3,490.01	0.20
	Uly.	3,302.75	0.18	3,369.57	0.26	3,429.39	0.32	3,489.99	0.14
24	Opt.	3,439.14	0.19	3,504.39	0.12	3,567.26	0.18	3,626.34	0.36
	Uly.	3,439.62	0.17	3,505.17	0.11	3,566.70	0.15	3,625.00	0.17
25	Opt.	3,574.71	0.17	3,640.71	0.19	3,703.67	0.26	3,763.14	0.52
	Uly.	3,574.87	0.18	3,639.06	0.19	3,703.94	0.10	3,763.12	0.24

Comparison of p -mode frequencies from optical observations (Opt.) and Ulysses electron channels E'1 to E'4 and DE1 to DE4 (Uly.). Each entry is a frequency, followed by its standard deviation. The entries marked '—' had no frequency estimates within the search tolerance. The Ulysses frequencies shown are the estimated frequencies divided by 1.00078.

The differences between the optical and Ulysses frequency estimates are often many standard deviations and the average error of the 90 matched modes is 1.03 μHz , 4.0 times the combined optical and Ulysses r.m.s. error; a histogram (not shown) of these differences is roughly trimodal with peaks near zero and $\pm 1 \mu\text{Hz}$. We speculate that this is a result of the strong 1.12 μHz (10.33 day sidereal period) line coupling through the nonlinearities of the electron propagation process or the logarithmic transform of the data used. For comparison, it is interesting to note that the differences between subsets of the optical frequency determinations in refs 11 and 42 have an average of 1.09 μHz , an r.m.s. discrepancy of 0.67 μHz and a similar trimodal histogram. There is some uncertainty (J. W. Harvey, personal communication) about clock accuracy in ref. 42.

To check further the hypothesis that p -mode frequencies are being observed, we replaced the frequency estimates with uniformly distributed random variates with the same range and repeated the grouping and 'Doppler' matching procedures. This averaged about 28 matches (maximum 38) with the 118 optical p -mode frequencies used, in contrast to the 90 matches obtained above.

Theoretical estimates of solar mode frequencies show a region between about 400 and 1,100 μHz where few modes are expected, see Fig. 4 in ref. 8. For comparison, Table 4 lists frequencies of groups below 2,100 μHz that initially had five or more frequen-

TABLE 4 Intermediate mode frequencies

Freq. (μHz)	s.d. (μHz)	t	R	N
1.11	0.15	1.11	13	13
124.14	0.17	2.30	4	5
132.02	0.34	6.18	6	8
146.40	0.24	2.58	6	8
163.56	0.27	3.93	6	8
212.22	0.17	2.38	4	5
262.46	0.31	2.68	4	6
271.62	0.37	3.15	5	6
445.26	0.24	2.42	4	6
458.73	0.20	7.50	5	7
522.30	0.46	5.18	4	5
529.45	0.40	4.07	4	6
558.00	0.19	4.86	4	6
674.36	0.19	4.25	4	5
728.63	0.24	6.04	4	6
788.71	0.22	9.46	4	6
805.36	0.31	4.92	4	6
862.49	0.24	4.01	4	6
869.79	0.15	7.30	4	6
909.76	0.27	2.04	5	6
1,034.68	0.25	3.83	4	6
1,095.69	0.28	8.76	4	5
1,107.27	0.28	1.61	4	5
1,231.59	0.28	2.15	4	5
1,294.10	0.19	6.08	4	6
1,340.61	0.27	3.59	5	6
1,465.50	0.18	1.00	5	6
1,488.85	0.24	2.91	5	6
1,563.73	0.24	5.04	4	5
1,753.13	0.24	5.62	4	6
1,759.28	0.22	3.57	4	6
2,003.02	0.33	4.17	5	6
2,094.16	0.59	1.96	4	5

Frequencies between those given in Table 2 and the p -modes of Table 3. These frequencies are all determined from the electron data used in Table 3 and have been divided by the same Doppler correction of 1.00078. Both the frequencies (Freq.) and standard deviations (s.d.) are robust estimates. The t -test statistic is the range within the group divided by the robust standard deviation estimate, and R is the number of samples remaining after trimming extreme values of the original N estimates.

cies. The frequencies in the table are trimmed means, and the standard deviations are from trimmed means of log-variances. The trimming was adjusted to give reasonable homogeneity between the average of the estimated jackknife variances and the variance estimated directly from the frequency estimates within each group; usually one or two extreme frequency estimates were discarded. Several of these frequencies coincide with peaks in the solar spectrum shown in Fig. 2 in ref. 43. Moreover, peaks in spectra of solar-wind data are commonly seen in the p -mode frequency range (for example, Figs 1a or 12a in ref. 44) but, as with our ion data, with insufficient resolution to resolve individual modes.

Magnetic and optical data

We took the frequencies estimated from the low-frequency particle data, Table 2, and tested interplanetary magnetic field (IMF) data for a periodic component within $\pm 0.005 \mu\text{Hz}$ of each of these frequencies. We used two long segments of OMNI⁴⁵ IMF data (Table 1) from August 1977 to December 1981 so there is no time-overlap between the magnetic and particle data sets. Because the frequencies are pre-specified, the problem is a classical binary hypothesis test and the significance levels of the F -statistic can be directly interpreted.

Of the 35 frequencies listed in Table 2, 13 are matched directly by frequencies in the IMF data within 0.005 μHz . In addition, because these data were acquired by near-Earth satellites, the longitude or ϕ dependence in equation (1) implies that one should expect mode splitting at multiples of one cycle per year (0.032 μHz). Allowing offsets of 1 and 2 cycles per year gives 28 matches. Further, the IMF data are from a pair of satellites. Switching the data source between the two causes an apparent additional line splitting near 0.897 μHz (about 12.9 days). Including this switching offset as well gives the 106 matches noted in Table 2, column 9. Finally, a histogram of the differences between the particle and magnetic data also shows splitting at 0.558 μHz , half the 1.12 μHz noted above, and similar to the Phobos optical data¹³. Including this further splitting gives 538 matches (Table 2, column 10) and all 35 particle frequencies are matched at least eight times. In these matches the minimum F -statistic from the magnetic data exceeds the 97% level.

Similarly, of the 14 low-frequency optical modes reported in ref. 1, six of the frequencies match the particle frequency groups

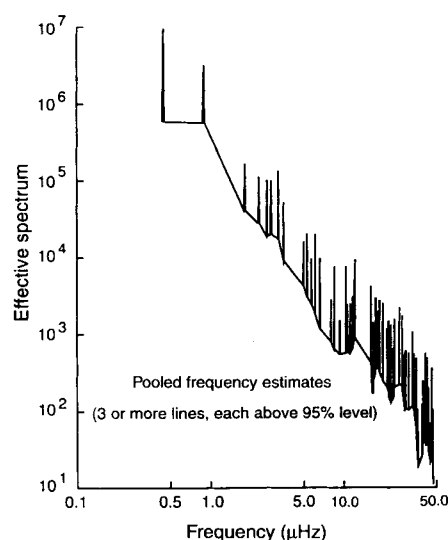


FIG. 3 Line spectrum in the g -mode frequency range derived from three or more frequency coincidences in the spectra of interplanetary particle fluxes.

containing three, or more, coincidences to within about 0.02 μHz .

Discussion and implications

Spectral analysis of low-energy charged particle fluxes measured in the interplanetary medium beyond the orbit of Earth reveals a wealth of periodic components. These are non-harmonic, yet strictly periodic in character. In the initial comparisons, they correspond well with those frequencies predicted by solar models, in these cases both the solar p modes and g modes. Modulation of the particle fluxes may be produced (locally or remotely) by the IMF. Thus, it is likely that the solar g modes are propagated into the interplanetary medium as fluctuations in the IMF, probably as large-scale Alfvén waves. Our results indicate that fluctuations in the IMF contain evidence of the spectrum of solar oscillations.

This propagation of the p - and g -mode frequencies into the interplanetary medium as fluctuations of the IMF has major implications for several important topics in interplanetary physics. These include the propagation of solar and Galactic particles, the access of Galactic and anomalous cosmic-ray particles to the heliosphere, and the modulation of the heliosphere termination shock. We note in particular that Jokipii and his colleagues^{9,46} have discussed a model for solar polar IMF conditions wherein the polar fields will have large transverse components and thus, at a given distance from the Sun, magnitudes large compared to those expected from simple extrapolations of spiral IMF models. They suggest that supergranulation (time-scales of the order of $(3-10) \times 10^4$ s, or 10–30 μHz) on the solar surface could cause their hypothesized field fluctuations. These granulations have a transverse velocity of the order of 0.5 km s^{-1}

on the solar surface. As our mode amplitudes in this frequency range are relatively large, it is possible that supergranulation is, in part, a surface manifestation of g modes.

The nature of the spectrum of the particle fluctuations is of importance in the context of interplanetary particle propagation. Figure 3 gives the power spectrum of the particle fluctuations in the frequency range 1–50 μHz . The individual lines correspond to frequencies of three or more coincidences, each above the 95% confidence level within a range of $\pm 0.1 \mu\text{Hz}$ around each frequency value. The ‘background’ of the spectrum, shown as the solid line connecting the lower edges of the spectral peaks, corresponds to the expected value of the average F -test for the cluster in noise. Essentially the entire spectrum is dominated by individual spectral lines, most in the frequency range corresponding to the solar g modes. This implies that charged particles are being scattered, and therefore diffused, in the interplanetary medium not only by a continuous spectrum of IMF fluctuations, but also, and perhaps more importantly, by discrete wave frequencies. Similarly, our observations of p modes in the electron data at 4.4 AU (Table 3) impose severe constraints on plasma turbulence, and fractal or chaotic models of the IMF.

The detrimental effects of solar particles on terrestrial systems are well documented; these range from radiation damage of spacecraft electronics and solar arrays, to production of induced voltages on telephone and power cables and corrosion on pipelines^{47,48}. For such problems, the implications of interplanetary particles having a nearly discrete spectrum are twofold: first, the ratio of peak to r.m.s. amplitudes is larger than if the spectrum was continuous; second, series with discrete spectra are, in principle, more predictable than series with a ‘similar’ continuous spectrum. $\square$

Received 23 December 1994; accepted 13 June 1995.

- Delache, P. & Scherrer, P. H. *Nature* **306**, 651–653 (1983).
- Gough, D. O. *Phil. Trans. R. Soc. A* **346**, 37–49 (1994).
- Bahcall, J. N. & Ulrich, R. K. *Rev. mod. Phys.* **60**, 297–372 (1988).
- Raghavan, R. S. *Science* **267**, 45–51 (1995).
- Bahcall, J. N. et al. *Nature* **375**, 29–34 (1995).
- Owens, A. J. & Jokipii, J. R. *J. geophys. Res.* **79**, 895–906 (1974).
- Burlaga, L. F. *J. geophys. Res.* **97**, 4283–4293 (1992).
- Christensen-Dalsgaard, J., Gough, D. O. & Toomre, J. *Science* **229**, 923–931 (1985).
- Jokipii, J. R. & Kóta, J. *Geophys. Res. Lett.* **16**, 1–4 (1989).
- Balogh, A. et al. *Science* **268**, 1007–1010 (1995).
- Libbrecht, K. G., Woodard, M. F. & Kaufman, J. M. *Astrophys. J. Suppl. Ser.* **74**, 1129–1149 (1990).
- Elsworth, Y. et al. *Astrophys. J.* **434**, 801–806 (1994).
- Toutain, T. & Fröhlich, C. *Astr. Astrophys.* **257**, 287–297 (1992).
- Gough, D. & Toomre, J. A. *Rev. Astr. Astrophys.* **29**, 627–684 (1991).
- Bahcall, J. N. & Kumar, P. *Astrophys. J.* **409**, L73–L76 (1993).
- Krimigis, S. M. et al. *Space Sci. Rev.* **21**, 329–354 (1977).
- Lanzerotti, L. J. et al. *Astr. Astrophys.* **92**, 349–364 (1992).
- Lanzerotti, L. J. et al. *Astrophys. J.* **380**, L93–L96 (1991).
- Dempster, A. P., Laird, N. M. & Rubin, D. B. *J. R. statist. Soc. B* **39**, 1–38 (1977).
- Thomson, D. J. *Bell Syst. tech. J.* **56**, 1769–1815; 1983–2005 (1977).
- Kleiner, B., Martin, R. D. & Thomson, D. J. *J. R. statist. Soc. B* **41**, 313–351 (1979).
- Martin, R. D. & Thomson, D. J. *Proc. IEEE* **70**, 1097–1115 (1982).
- Chave, A. D. & Thomson, D. J. *J. geophys. Res.* **94**, 14215–14225 (1989).
- Thomson, D. J. *Phil. Trans. R. Soc. A* **330**, 601–616 (1990).
- Lindberg, C. R. thesis, Univ. California, San Diego (1986).
- Park, J., Lindberg, C. R. & Thomson, D. J. *Geophys. J. R. astr. Soc.* **91**, 755–794 (1987).
- Lindberg, C. R. & Thomson, D. J. *J. geophys. Res.* **95**, 19785–19788 (1990).
- Thomson, D. J. *Proc. IEEE* **70**, 1055–1096 (1982).
- Thomson, D. J. *Phil. Trans. R. Soc. A* **332**, 539–597 (1990).
- Percival, D. B. & Walden, A. T. *Spectral Analysis for Physical Applications; Multitaper and Conventional Univariate Techniques* (Cambridge Univ. Press, 1993).

- Rife, D. C. & Boorstyn, R. R. *Bell Syst. tech. J.* **55**, 1389–1410 (1976).
- Thomson, D. J. & Chave, A. D. in *Advances in Spectrum Analysis* Ch. 2 (ed. Haykin, S.) (Prentice-Hall, Englewood Cliffs, NJ, 1990).
- Pearson, E. S. & Hartley, H. O. *Biometrika Tables for Statisticians* (Cambridge Univ. Press, 1972).
- Snyder, D. L. *Random Point Processes* (Wiley, New York, 1975).
- Kendall, M. G. & Stuart, A. *The Advanced Theory of Statistics* Vol. 1 (Hafner, New York, 1963).
- Hasan, T. in *Handbook of Statistics* Vol. 3 (eds Brillinger, D. R. & Krishnaiah, P. R.) Ch. 7 125–156 (Elsevier, Amsterdam, 1983).
- Middleton, D. *Statistical Communication Theory* (McGraw Hill, New York, 1960).
- Scherrer, P. H. & Wilcox, J. M. *Solar Phys.* **82**, 37–42 (1983).
- Lanzerotti, L. J. & MacLennan, C. G. *Nature* **275**, 113–114 (1979).
- Toth, N. *Nature* **270**, 159–160 (1977).
- Vladimirov, B. M., Bobova, V. P., Bondarenko, N. M. & Veretennikova, V. K. *Solar Phys.* **82**, 451–455 (1983).
- Grec, G., Fossat, E. & Pomerantz, M. A. *Solar Phys.* **82**, 55–66 (1983).
- Brown, T. M., Gilliland, R. L., Noyes, R. W. & Ramsey, L. W. *Astrophys. J.* **368**, 599–609 (1991).
- Goldstein, M. L., Roberts, D. A. & Fitch, C. A. *J. geophys. Res.* **99**, 11519–11538 (1994).
- Couzens, D. A. & King, J. H. *Interplanetary Medium Data Book—Suppl. 3A*, 1977–85 (NSSDC/WDC-A, Greenbelt, MD, 1986); also on *Selected Geomagnetic and other Solar-Terrestrial Physics Data of NOAA and NASA NGDC-01 CD-ROM* (NOAA, Boulder, CO, 1987).
- Jokipii, J. R. *Adv. Space Res.* **9**, 105–119 (1989).
- Lanzerotti, L. J., Maurer, D. W., Sauer, H. H. & Zwicky, R. D. *J. Spacecraft Rockets* **28**, 614–616 (1991).
- Lanzerotti, L. J., Medford, L. V., MacLennan, C. G. & Thomson, D. J. *AT&T Tech. J.* **74**, 73–84 (1995).

ACKNOWLEDGEMENTS. We thank the Ulysses HI-SCALE team and the ESA and NASA Ulysses projects for their continuing efforts that have made this work possible. We thank D. Gough for numerous comments, scrutiny of the p -mode data, and general advice on helioseismology. J. Harvey, T. M. Brown and particularly J. Lagarias made helpful comments.

Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene *CBL2*

C. Jones^{*†}, L. Penny[‡], T. Mattina[¶], S. Yu[§], E. Baker[§], L. Voullaire^{**}, W. Y. Langdon^{††}, G. R. Sutherland[§], R. I. Richards[§] & A. Tunnacliffe^{***‡‡}

[†] Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

[‡] Division of Medical Genetics, Department of Medicine, University of California San Diego, La Jolla, California 92093, USA

[¶] Clinica Pediatrica, I Università di Catania, viale Andrea Doria, 6, Catania, Sicily, Italy

[§] Centre for Medical Genetics, Department of Cytogenetics and Molecular Genetics, Adelaide Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia

^{**} Murdoch Institute, 10th Floor Royal Children's Hospital, Flemington Road, Melbourne, Victoria 3052, Australia

^{††} Department of Biochemistry, University of Western Australia, Nedlands, West Australia 6009, Australia

^{‡‡} Quadrant Research Foundation, Maris Lane, Trumpington, Cambridge CB2 2SY, UK

The fragile site *FRA11B* has been localized to the p(CCG)_n repeat of the *CBL2* proto-oncogene. A proportion of Jacobsen (11q⁻) syndrome patients inherited a chromosome carrying a *CBL2* p(CCG)_n expansion, which was truncated close to *FRA11B*. These results have broad implications for the role of p(CCG)_n repeat expansion in the aetiology of genetic disease involving chromosome rearrangements.

FOLATE-sensitive fragile sites are inherited chromosomal aberrations that appear as poorly condensed regions of metaphase chromosomes¹ and are susceptible to breakage under specific experimental conditions^{1,2}. So far, however, there has been no direct evidence of chromosomal breakage occurring *in vivo* in carriers of fragile sites.

Of the folate-sensitive fragile sites that have been characterized at the molecular level (*FRAXA*, *FRAXE*, *FRAXF* and *FRA16A*)³⁻⁹, only two are linked to a clinical phenotype: *FRAXA* is associated with fragile X syndrome^{10,11}, the second most common cause of mental retardation after Down's syndrome¹², and *FRAXE* with a rare form of mild mental retardation^{7,13,14}. Because of the clinical importance of fragile X syndrome, *FRAXA* was the first fragile site to be characterized. A gene associated with fragile X syndrome (*FMRI*) has been identified⁴ and shown to contain specific molecular anomalies in affected individuals^{3,15,17}. The molecular basis of *FRAXA* is the extensive expansion of a CCG-trinucleotide repeat in the 5' untranslated region of *FMRI*⁴⁻⁶ and associated methylation of an adjacent CpG island^{16,17}. In most cases, the clinical presentation of fragile X syndrome is caused by a reduction in the transcription of *FMRI*^{18,19} as a result of this *FRAXA*-induced methylation.

More recently, other folate-sensitive fragile sites have also been found to be associated with p(CCG)_n repeat expansion and changes in regional methylation status⁷⁻⁹. Characterization of the autosomal fragile site *FRA16A* provided evidence that this methylation is a consequence of trinucleotide expansion, rather than the manifestation of an intrinsic imprinting system (such as X-inactivation or parental imprinting)⁹.

Database searches have identified numerous human genes which have p(CCG)_n repeats located in their 5' untranslated regions²⁰, but only one of these, *CBL2*²¹, is located in the vicinity^{22,23} of a rare folate-sensitive fragile site, *FRA11B*²⁴. The cloning of the *CBL2* gene in yeast artificial chromosomes (YACs) and cosmids has recently allowed the mapping of the p(CCG)_n repeat to within a 100 kb region of 11q23.3 that also

contains *FRA11B*²⁵. We report here the localization of *FRA11B* to the *CBL2* p(CCG)_n repeat and demonstrate an association with the chromosome deletion characteristic of Jacobsen (11q⁻) syndrome.

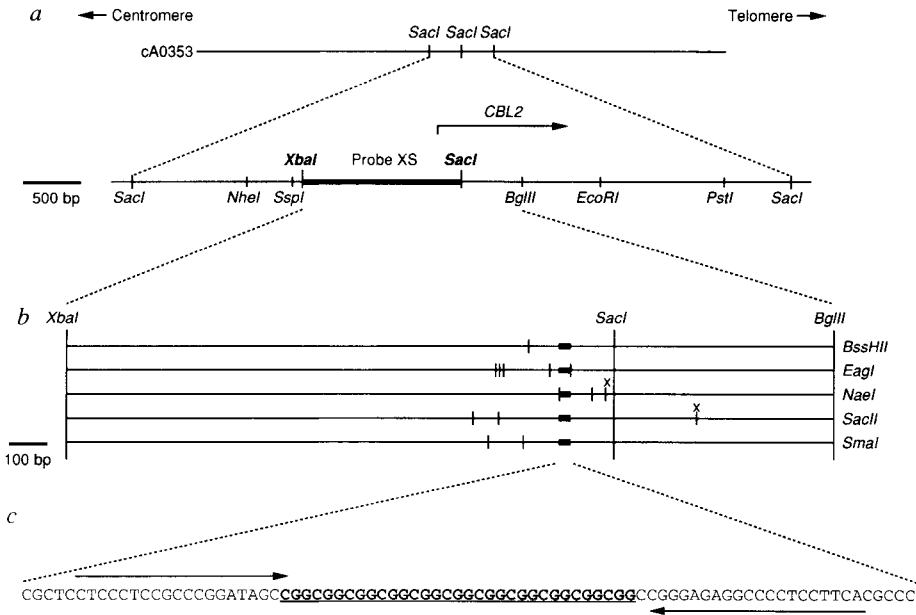
Localization of *FRA11B*

We have previously described a cosmid, cA0353, which contains the 5' end of the *CBL2* gene and includes the p(CCG)_n repeat²⁵. Fluorescence *in situ* hybridization (FISH) on metaphase spreads from *FRA11B*-expressing cells shows that the signal from cA0353 is localized to the fragile site and flanking sequences: of 24 *FRA11B* chromosomes counted, the cA0353 signal was proximal in 14, central in three, distal in two and both proximal and distal in five cases. This demonstrates that sequences within cA0353 either flank the fragile site or are very close to it and is therefore consistent with the fragile site being at the *CBL2* trinucleotide repeat.

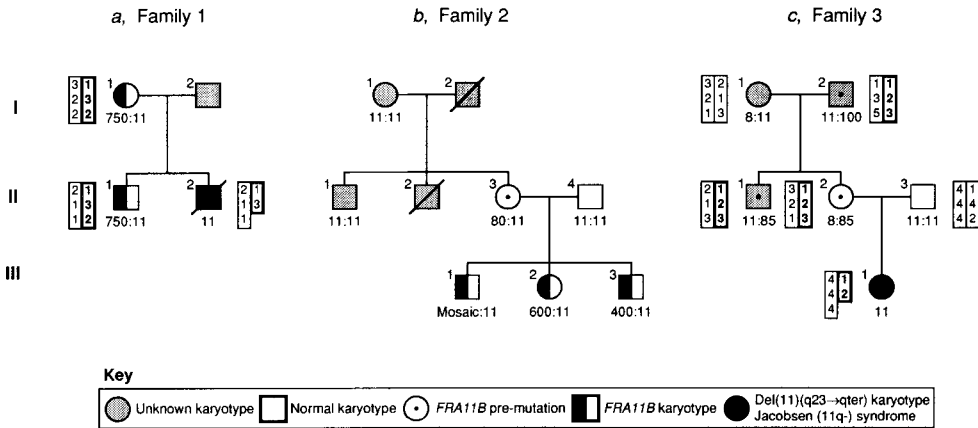
A restriction map of the part of cosmid cA0353 containing the 5' end of *CBL2* is shown in Fig. 1a and b. The DNA sequence of the 1.5 kb *XbaI/SacI* fragment that includes the *CBL2* p(CCG)_n repeat was determined from subclones of this cosmid (data not shown; accession number X82654) and primers were synthesized for use in polymerase chain reaction (PCR) analysis of the repeat (Fig. 1c) in *FRA11B*-expressing and non-expressing chromosomes. The majority of non-expressing chromosomes (>70% of 130 chromosomes) are found to have a copy number of *n* = 11 for the p(CCG)_n repeat. This low level of heterozygosity, together with the small number of *FRA11B* families available, prevented the demonstration of inheritance of a null allele in a *FRA11B* individual, associated with an inability to amplify expanded p(CCG)_n repeats by PCR. Instead, the stability of the *CBL2* p(CCG)_n repeat was examined by Southern analysis using as a probe the 1.5 kb *XbaI/SacI* restriction fragment (probe XS; Fig. 1a), which includes the trinucleotide repeat.

The mother in family 1 (Fig. 2a) expresses *FRA11B* and has transmitted the fragile site to at least one of her sons²⁶. Following digestion with the enzymes *SspI* and *SacI*, DNA from non-expressing individuals gives rise to a 1.5 kb band that hybridizes to the XS probe. However, the son (II-1) of

* To whom correspondence should be addressed.



Family 2 (Fig. 2b) contains two individuals (III-1 and III-3) who were reported, after cytogenetic analysis, to express *FRA11B*. Both individuals were found to have an additional band or bands of hybridization with the XS probe (Fig. 3b). The additional band seen in individual III-3 represents a size increase of 1.2 kb and therefore corresponds to a p(CCG)_n.



polymorphic microsatellite (AC)_n repeat primer pairs (Généthon, Evry, France)⁴⁵ as described³⁰. In descending order in the boxes: D11S1356, D11S1341 and D11S925 for family 1; D11S1347, D11S1341 and D11S925 for family 3. In addition, loss of heterozygosity analysis of Jacobsen (11q⁻) patients was confirmed using the additional primer pairs: D11S939, D11S924, D11S1336, D11S1345, D11S1353, D11S1328, D11S936, D11S1316, D11S933, D11S934, D11S1351, D11S912, D11S910, D11S1320, D11S969, D11S1309, D11S968.

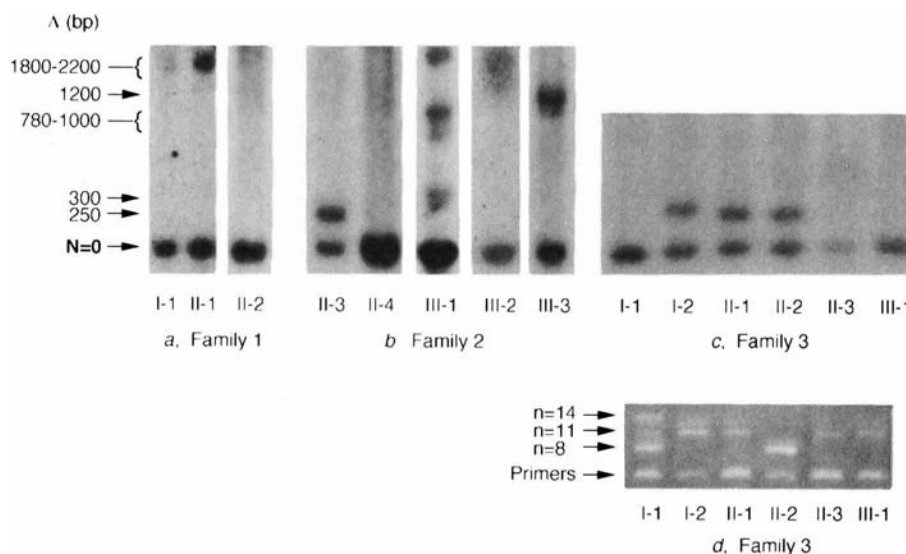


FIG. 3 a–c, Southern analysis of *FRA11B* families. Chromosomal DNA from *FRA11B* families was digested with the restriction enzymes *Xba*I and *Bgl*II, or *Sac*I and *Ssp*I, and subjected to Southern blot analysis using the XS probe. The normal band is indicated by the letter N. Size increases (bp) corresponding to *CBL2* p(CCG)_n repeat amplification are shown at the left of the figure (Δ). d, PCR analysis of *CBL2* p(CCG)_n repeat lengths (n) in members of family 3.

METHODS. a–c, Genomic DNA was prepared from fresh blood lymphocytes by standard procedures. Restriction digestions of 10 μ g of DNA were performed with 100 units of enzyme in a 100- μ l volume overnight. DNA was electrophoresed on 1% agarose gels and blotted onto Zeta-probe GT nylon membranes (Bio-Rad) by alkaline capillary transfer using standard techniques⁴⁶. Probe XS was excised from a pUC clone of the 3-kb *Sac*I fragment, separated on 1% LMP agarose and purified by treating with agarase (Sigma) followed by phenol extraction and ethanol precipitation. The probe was labelled by a standard procedure using random hexamers (Pharmacia), Klenow polymerase (Boehringer Mannheim) and 20 μ Ci of α -³²P-dCTP (Amersham). Prehybridization (with 0.5 mg ml⁻¹ sonicated salmon sperm DNA) and hybridization with probe

XS were performed in a standard dextran sulphate/Denhardt's hybridization buffer⁴⁶ at 65 °C for 2 h and 16 h respectively. Filters were washed for 30 min in 2 $\times$ SSC/0.1% SDS at room temperature, followed by washing for 30 min in 0.2 $\times$ SSC/0.1% SDS at 65 °C, before exposure to X-OMAT AR X-ray film (Kodak). Images of autoradiograms were processed using Adobe 'Photoshop' software, only by adjusting contrast and brightness of whole images, in order to reveal appropriate bands more clearly. d, PCR reactions were performed on 50 ng genomic DNA (predenatured in 0.4 M NaOH; 0.4 mM EDTA at 37 °C for 30 min, neutralized in 0.3 M sodium acetate and ethanol precipitated); using 1U Gibco-BRL *Taq* DNA polymerase in the supplied buffer; plus 1.5 mM MgCl₂; 250 mM each dATP, dCTP, dTTP, 60 mM dGTP and 190 mM dITP; 2 mM each primer (Fig. 1c); in a 10- μ l reaction volume for 1 cycle of 98 °C, 1 min; 70 °C, 5 min; 40 cycles of 94 °C, 1 min; 70 °C, 3 min. PCR products were separated on a 6% neutral polyacrylamide gel and visualized using ethidium bromide. Allele *n* = 14 in individual I-1 of family 3 is most probably the result of somatic variation, as indicated by germ-line transmission of alleles *n* = 11 and *n* = 8 to individuals II-1 and II-2 respectively.

repeat amplification with a copy number of approximately *n* = 400. Interestingly, individual III-1 shows several additional bands with the XS probe and is probably mosaic for p(CCG)_n repeat expansion with several repeat lengths of between *n* = 80 and *n* > 1000. The sister (III-2) of the two *FRA11B*-expressing boys was also shown to carry an expanded fragment corresponding to a repeat expansion to *n* = 600–650, but was not previously known to carry the fragile site. It was predicted that she would express *FRA11B* and subsequent cytogenetic analysis showed this to be the case.

The father of the *FRA11B* individuals in family 2 (II-4) showed no evidence of expansion at the *CBL2* trinucleotide, whereas the cytogenetically normal mother (II-3) had an additional band corresponding to a p(CCG)_n copy number of approximately *n* = 80 (Fig. 3b). It is likely that the *FRA11B*-carrying offspring inherited the maternal chromosome containing this trinucleotide expansion and that this has undergone further amplification to give rise to the fragile site. The p(CCG)_n repeat length in II-3 is close to the pre-mutation range in fragile X syndrome, where the individual is normal but whose offspring are predisposed to further expansion of the *FMR1* p(CCG)_n repeat, expression of *FRAXA* and hence fragile X syndrome^{6,27}. Three individuals in family 3 (I-2, II-1 and II-2; Fig. 2c) have similar-sized repeats (*n* = 100, 85 and 85 respectively; Fig. 3c) of whom individual II-2 is also known to have a normal karyotype. Interestingly, the *n* = 100 allele of individual I-2 has been reduced in size upon transmission to both of his offspring, to *n* = 85. It

is probable, therefore, that a low-level expansion (*n* $\leq$ 100) of the *CBL2* trinucleotide represents a pre-mutation for *FRA11B*.

In total, of six *FRA11B* families studied, all nine of the known *FRA11B*-expressing individuals were found to have expansions in the *CBL2* p(CCG)_n repeat of *n* > 100 compared with non-expressing individuals examined, and one other individual was confirmed to express *FRA11B* following prediction from Southern analysis. We have identified families in which repeat expansion and reduction have occurred, families in which an expanded repeat was stably transmitted, and individuals carrying pre-mutation and mosaic repeat lengths. In summary, Southern analysis presents strong evidence that expansion of the *CBL2* p(CCG)_n trinucleotide is responsible for the expression of *FRA11B*.

Further evidence that the additional bands seen in Southern analysis are caused by expansion of the *CBL2* trinucleotide was obtained from PCR analysis of the p(CCG)_n repeat (Fig. 1c) in individuals of family 3 (Figs 2c and 3d). Although we have been unable to demonstrate the inheritance of a null allele corresponding to the full repeat expansion of *FRA11B*-expressing individuals, this was possible for the pre-mutation alleles (*n* = 85–100) present in members of family 3. PCR analysis of this family suggests that individual I-2 is homozygous *n* = 11 (as the *n* = 100 allele is not amenable to PCR) whereas his wife (I-1) shows somatic variation and is heterozygous for three alleles, *n* = 8/11/14 (Fig. 3d). If individual I-2 were actually homozygous *n* = 11, the daughter of this couple (II-2) should carry

at least one $n=11$ allele. That she does not, but seems to be homozygous $n=8$ by PCR (Fig. 3d), shows the presence of a null allele in this family and is considered proof that the *CBL2* p(CCG) $_n$ repeat is expanded as predicted by the Southern analysis.

Methylation analysis in *FRA11B* chromosomes

Expansion of p(CCG) $_n$ repeats at other fragile sites is associated with hypermethylation of adjacent CpG islands when repeat length exceeds certain limits^{7,9,16,17}. Methylation of the CpG island adjacent to the *FRA11B* p(CCG) $_n$ repeat causes a loss of *FMRI* expression^{18,19}, giving rise to the clinical phenotype associated with fragile X syndrome. Genomic sequence analysis of the region surrounding the *CBL2* p(CCG) $_n$ repeat identifies a CpG island immediately adjacent to the repeat which contains several recognition sites for rare-cutting restriction enzymes typical of such regions. To determine the methylation status of *FRA11B* chromosomes, we performed Southern analysis on DNA from an EBV-transformed cell line, MB, derived from individual III-3 of family 2. DNA from cell line MB was cut with the enzymes *Xba*I/*Bgl*II and one of the methylation-sensitive enzymes *Sac*II or *Sma*I (Fig. 4). The methylation status of restriction sites for the enzymes *Bss*HII, *Eag*I and *Nae*I was also established (data not shown). The locations of the restriction sites for these enzymes within the *CBL2* CpG island are shown in Fig. 1b. DNA from several individuals who do not express *FRA11B* is cut by all of these enzymes (Figs 4a, b and data not shown) at the sites indicated in Fig. 1b. But digestion of DNA from cell line MB shows that all of the restriction sites 5' to the expanded trinucleotide repeat are methylated, as none is cut by the relevant enzyme (Fig. 4c and data not shown). In contrast, DNA from cell line MB is cut at the *Sac*II site (Fig. 4c) and at least one *Nae*I site (data not shown) 3' to the expanded repeat, indicating that these sites are not methylated. Similar patterns of methylation were also demonstrated in both EBV-transformed cell lines and fresh blood lymphocytes of other *FRA11B*-expressing individuals (data not shown). Thus, like other fragile sites so far characterized, a CpG island adjacent to *FRA11B* is found to be methylated when p(CCG) $_n$ repeat length exceeds certain limits.

Association of *FRA11B* with Jacobsen syndrome

Jacobsen (11q⁻) syndrome is the clinical presentation of the loss of part of the long arm of chromosome 11, typically from band 11q23→qter, and is recognized by specific dysmorphic features and moderate mental retardation^{28,29}. Voullaire *et al.*²⁶ reported that individual II-2 from family 1 (Fig. 2a) had a deletion del(11)(q23→qter) of one chromosome 11 homologue and had signs of Jacobsen syndrome. We have recently mapped the chromosome 11 deletion breakpoint in this Jacobsen patient to the same 100 kb region of 11q23.3 that includes *FRA11B*²⁵. In addition,

a separate study shows clustering of chromosome 11 deletion breakpoints in five other Jacobsen patients between markers *D11S1356* and *D11S925* (ref. 30), a region of approximately 5 cM which encompasses *FRA11B*. This mapping information suggests a relationship between *FRA11B* and at least a proportion of Jacobsen syndrome deletions and we therefore attempted to examine this relationship more closely.

Figure 3a shows that the Jacobsen patient of family 1 (II-2) does not contain an expanded fragment that is typical of *FRA11B* and that is present in both the mother (I-1) and sibling (II-1). However, the haplotype data presented in Fig. 2a allow us to identify which maternal chromosome carries the fragile site (haplotype 1,3,2 in Fig. 2a). In addition, haplotype analysis also indicates that this same maternal chromosome (that is, the *FRA11B* chromosome) has been inherited by the Jacobsen patient (II-2) and has suffered deletion. Therefore these results clearly demonstrate that the 11q⁻ chromosome of the Jacobsen patient derives from the maternal *FRA11B* chromosome.

Southern analysis using DNA from the families of four further Jacobsen patients also detected a *CBL2* p(CCG) $_n$ repeat expansion in one other case (family 3; Figs 2c and 3c). Cytogenetic analysis did not reveal *FRA11B* expression in either of the parents of this family. However, the mother (II-2) was found to carry an expansion of the *CBL2* p(CCG) $_n$ repeat to $n=85$ (Fig. 3c) which corresponds to a *FRA11B* pre-mutation (see above). Haplotype analysis of this family indicates that the deleted chromosome of the Jacobsen patient (III-1) derives from the maternal grandfather (I-2). Southern analysis demonstrates that the chromosome carrying the expanded *CBL2* trinucleotide also originates from this individual. Together, these observations show that, in this family also, it is the chromosome carrying the expanded *CBL2* trinucleotide that has suffered deletion. Therefore an association between *CBL2* p(CCG) $_n$ repeat expansion and chromosome truncation has been found in two out of five Jacobsen syndrome patients.

Jacobsen chromosome truncation near *FRA11B*

Southern analysis of DNA from the Jacobsen syndrome patients using probe XS does not detect rearrangements at the *CBL2* p(CCG) $_n$ (Fig. 3a and c), suggesting that these deletions may have occurred some distance from *FRA11B*. We were therefore interested to localize the Jacobsen deletion breakpoints more accurately.

FISH on metaphase spreads from cells derived from two Jacobsen patients shows that cosmid cA0353 hybridizes to both the normal and the deleted chromosome 11 in more than 20 metaphase spreads counted for each individual. But detailed analysis of one of these patients showed that, whereas probe XS and the 3 kb *Sac*I fragment distal to *FRA11B* both hybridize to the normal chromosome 11, neither probe hybridizes to the 11q⁻ chromosome. The chromosome deletion breakpoint was there-

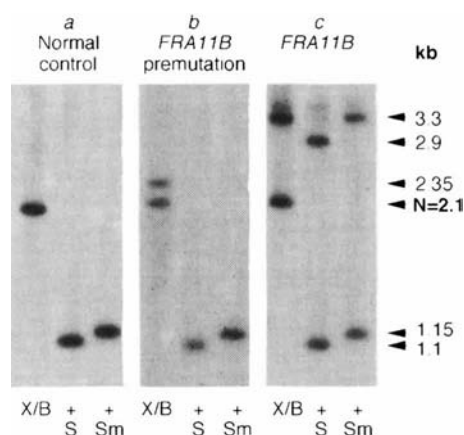


FIG. 4 Methylation analysis of the *CBL2* CpG island in a, normal; b, *FRA11B* pre-mutation; and c, *FRA11B*-expressing individuals. Chromosomal DNA was cut with the restriction enzymes *Xba*I and *Bgl*II, plus the methylation-sensitive restriction enzymes indicated, and subjected to Southern blot analysis using the XS probe (see Fig. 3 methods). The normal 2.1-kb band is indicated by the letter N and the sizes of other bands (in kb) are shown at the right of the figure. B = *Bgl*II; S = *Sac*II; Sm = *Sma*I; X = *Xba*I.

fore not located directly at the *CBL2* p(CCG)_n repeat, but must be within the 20 kb of cA0353 proximal to *FRA11B*. This is consistent with *in vitro* observations of *FRA11B*-induced translocations, where chromosome breakpoints map within several kilobases of the *FMR1* trinucleotide repeat⁴.

Discussion

This work is the first to localize a fragile site within a previously described gene, highlighting the possibility that other p(CCG)_n repeats contained within known genes may also be associated with folate-sensitive fragile site expression. Other than their direct association, the precise roles of repeat expansion and methylation in causing fragile site expression are still not understood. The *CBL2* p(CCG)_n repeat is not highly polymorphic in the general population, indicating that polymorphism is not a prerequisite for possible repeat expansion. In the light of this observation, it will be interesting to test the hypothesis that expansion of any p(CCG)_n repeat adjacent to a CpG island could give rise to a fragile site.

Given the very low incidence of both *FRA11B* and Jacobsen syndrome, an association between them is extremely unlikely to be due to chance. Hence these results are the first to demonstrate a direct link between a fragile site and chromosome breakage *in vivo*. Chromosomal deletions and rearrangements associated with genetic disease were previously believed to occur *de novo*, with little or no impact of genetic background. The demonstration of an inherited component in the development of at least a proportion of cases of Jacobsen syndrome challenges that dogma. Furthermore, the implication of a fragile site in the generation of the deletion provides insight into the mechanism of the truncation event.

The parents of three Jacobsen patients examined do not seem to express *FRA11B*, and it is possible that other uncharacterized fragile sites are responsible for these breaks. Seven p(CCG)_n repeats have been found within a 2-Mb region of chromosome 4 (ref. 31) and three folate-sensitive fragile sites have been mapped within 1–2 Mb on the X chromosome³². Hence it is possible that more than one fragile site could map within the Jacobsen breakpoint cluster region. Less probable, given that there is no evidence from other fragile sites, is that these patients could have suffered *de novo* expansions of the *CBL2* p(CCG)_n repeat that led to chromosome breakage.

It is clear that inheritance of *FRA11B* does not necessarily lead to chromosome deletion and consequent Jacobsen syndrome. However, a proportion of fragile chromosomes within a *FRA11B* carrier may still undergo truncation. For example, individual III-1 of family 2 shows truncation in 0.5% of metaphase chromosomes even without fragile site induction (T.M., unpublished data). Therefore, the proportion of 11q⁻ cells within an individual, which depends on the stage of development at which the deletion occurs, may be critical for the manifestation of a clinical phenotype. Other influences, such as environmental factors, may affect the likelihood of chromosome breakage occurring at fragile sites *in vivo*, as they do *in vitro*^{3,5,33}.

The role of fragile sites in oncogenesis has been the subject of much controversy and the localization of *FRA11B* to within *CBL2*, a known proto-oncogene, is intriguing. In fragile X syndrome, hypermethylation following p(CCG)_n repeat expansion^{16,17} results in the loss of transcription of the *FMR1* gene^{18,19}, which is responsible for the clinical phenotype. *CBL2* is known to be involved in a receptor tyrosine kinase signal transduction pathway^{34,35}, but if loss of *CBL2* transcription occurs at a *FRA11B*-expressing allele, it is not obviously linked to a phenotype. Considering the role of *CBL2* as a dominant oncogene, it also seems unlikely that the loss of one allele would be of any oncogenic consequence. In contrast, the potential role of a fragile site located in the 5' end of a tumour suppressor gene would be more apparent. Loss of transcription of such a hypothetical gene as a result of fragile site expression would represent the first 'hit' in the tumorigenic process.

A more significant influence of fragile site expression in tumorigenesis might be to increase the incidence of chromosome rearrangements or deletions during tumour progression. Deletions at specific chromosomal loci are key secondary events in the development of tumours that require loss of expression of a tumour suppressor gene³⁶, whereas rearrangements such as translocations are common primary events in the development of tumours that require activation of an oncogene³⁷. Specifically, mutations in DNA repair enzymes^{38–41} that have been implicated in microsatellite and trinucleotide repeat instability in some tumours^{42–44} could give rise to an expansion of p(CCG)_n repeat sequences throughout the genome and thus be indirectly responsible for the chromosome rearrangements necessary for tumour progression. □

- Sutherland, G. R. & Hecht, F. *Fragile Sites On Human Chromosomes* (Oxford Univ. Press, Oxford, 1985).
- Warren, S. T., Zhang, F., Licameli, G. R. & Peters, J. F. *Science* **237**, 420–423 (1987).
- Yu, S. et al. *Science* **252**, 1179–1182 (1991).
- Verkerk, A. J. M. H. et al. *Cell* **65**, 905–914 (1991).
- Kremer, E. J. et al. *Science* **252**, 1711–1714 (1991).
- Fu, Y.-H. et al. *Cell* **67**, 1047–1058 (1991).
- Knight, S. J. et al. *Cell* **74**, 127–134 (1993).
- Parrish, J. E. et al. *Nature Genet.* **8**, 229–235 (1994).
- Nancarrow, J. K. et al. *Science* **264**, 1938–1941 (1994).
- Giraud, F., Aymé, S., Mattei, J. F. & Mattei, M. G. *Hum. Genet.* **34**, 125–136 (1976).
- Harvey, J., Judge, C. & Wiener, S. J. *med. Genet.* **14**, 46–50 (1977).
- Brown, W. T. & Jenkins, E. C. *Molec. Genet. Med.* **2**, 39–66 (1992).
- Flynn, G. A. et al. *J. med. Genet.* **30**, 97–100 (1993).
- Mulley, J. C. et al. *J. med. Genet.* **32**, 162–169 (1995).
- Bell, M. V. et al. *Cell* **64**, 861–866 (1991).
- Oberle, I. et al. *Science* **252**, 1097–1102 (1991).
- Vincent, A. et al. *Nature* **349**, 624–626 (1991).
- Pieretti, M. et al. *Cell* **66**, 817–822 (1991).
- Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P. & Mandel, J. L. *Nature Genet.* **4**, 335–340 (1993).
- Richards, R. I. & Sutherland, G. R. *Cell* **70**, 709–712 (1992).
- Blake, T. J., Shapiro, M., Morse, H. C. & Langdon, W. Y. *Oncogene* **6**, 653–657 (1991).
- Savage, P. D. et al. *Cytogenet. cell. Genet.* **56**, 112–115 (1991).
- Tunnacliffe, A. & McGuire, R. S. *Genomics* **8**, 447–453 (1990).
- Sutherland, G. R., Jacky, P. B., Baker, E. & Manuel, A. *Am. J. hum. Genet.* **35**, 432–437 (1983).
- Jones, C. et al. *Hum. molec. Genet.* **3**, 2123–2130 (1994).

- Voulaire, L. E., Webb, G. C. & Leversha, M. A. *Hum. Genet.* **76**, 202–204 (1987).
- Rousseau, F. et al. *New Engl. J. Med.* **325**, 1673–1681 (1991).
- Jacobsen, P. et al. *Hum. Hered.* **23**, 568–585 (1973).
- Schinz, A., auf der Maur, P. & Moser, H. J. *med. Genet.* **14**, 438–444 (1977).
- Penny, L. A. et al. *Am. J. hum. Genet.* **56**, 676–683 (1995).
- Hummerich, H. et al. *Hum. molec. Genet.* **3**, 73–78 (1994).
- Willems, P. J. *Nature Genet.* **8**, 213–215 (1994).
- Warren, S. T. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3856–3860 (1990).
- Donovan, J. A., Wang, R. L., Langdon, W. Y. & Samelson, L. E. *J. biol. Chem.* **269**, 22921–22924 (1994).
- Andonou, C. E., Thien, C. B. F. & Langdon, W. Y. *EMBO J.* **13**, 4515–4523 (1994).
- Knudson, A. G. *Adv. Cancer Res.* **17**, 317–324 (1973).
- Yunis, J. J. *Science* **221**, 227–236 (1983).
- Fishel, R. et al. *Cell* **75**, 1027–1038 (1993).
- Leach, F. S. et al. *Cell* **75**, 1215–1225 (1993).
- Bronner, C. E. et al. *Nature* **368**, 258–261 (1994).
- Papadopoulos, N. et al. *Science* **263**, 1625–1629 (1994).
- Thibodeau, S. N., Bren, G. & Schaid, D. *Science* **260**, 816–819 (1993).
- Peltomäki, P. et al. *Cancer Res.* **53**, 5853–5855 (1993).
- Ma, L. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9871–9875 (1994).
- Gyapay, G. et al. *Nat. Genet.* **7**, 246–339 (1994).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989).

ACKNOWLEDGEMENTS. This work was supported by grants from the Medical Research Council to A.T.; the National Health and Medical Research Council (Australia) to R.I.R. and G.R.S.; and an International Research Scholars award from the Howard Hughes Medical Institute to G.R.S. We thank B. Ponder (Cambridge) for providing support and facilities to C.J. and A.T.

Interacting elliptical galaxies as hosts of intermediate-redshift quasars

M. J. Disney*, P. J. Boyce*, J. C. Blades†, A. Boksenberg‡, P. Crane§, J. M. Deharveng||, F. Macchetto¶, C. D. Mackay¶, W. B. Sparks† & S. Phillips**

* Department of Physics and Astronomy, University of Wales College of Cardiff, PO Box 913, Cardiff CF2 3YB, UK

† Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

‡ Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ, UK

§ European Southern Observatory, Karl Schwarzschild Strasse 2, D-8046 Garching, Germany

|| Laboratoire d'Astronomie Spatiale du CNRS, Traverse du Siphon, Les Trois Lucs, F-13012 Marseille, France

¶ Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

QUASARS are the most luminous objects in the Universe. It has been speculated that they are the visible evidence for accretion of gas onto supermassive black holes that reside at the centres of host galaxies. Direct observational confirmation that quasars reside in the centres of galaxies has been hard to obtain, because atmospheric turbulence usually scatters the quasar light sufficiently to swamp the signal from the fainter surrounding galaxy. Despite the difficulties, however, many attempts have been made to observe the host galaxies¹⁻¹⁷, although the results have not been definitive¹⁸. Here we report observations of four quasars, made with the refurbished Hubble Space Telescope. In all four cases the quasars reside in luminous elliptical galaxies with very close companions. This is in contrast to recent work^{19,20} in which host galaxies were not observed (in a sample of quasars that has one in common with ours). The elliptical galaxies are featureless, but the presence of close companions is suggestive of continuing interactions.

We present observations of the first four quasars from a larger sample of 20 that are being observed. The overall sample was chosen to include examples of all the main classes of quasars (radio-loud, radio-quiet, X-ray-selected and IRAS quasars). All have V-band magnitude $M_V \leq -23.0$ ($H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0$ used throughout), making them unequivocal quasars by the criterion used by Veron-Cetty and Veron²¹. The quasars were observed in late 1994 with the Planetary Camera (PC) within the Wide Field/Planetary Camera 2 (WFPC2)²² on the Hubble Space Telescope (HST). The PC has a field of 800×800 CCD pixels of size 0.046 arcsec per pixel. Each observation consisted of three 600-s integrations through the F702W filter which corresponds roughly to the R band-pass. This camera/filter combination was chosen for its high sensitivity to potential galaxy light and because it has sufficient resolution and dynamic range to avoid saturation in those critical areas of the image where quasar and galaxy can be most easily distinguished. Initial data processing involved the HST standard data reduction procedures whereas subsequent reductions were done on STARLINK software, making use of STSDAS software packages supplied by the Space Telescope Science Institute. After bias-subtraction, flat-fielding and calibration, the individual frames were median combined (to remove cosmic-ray events) to yield the 'raw images'.

A great deal of information about a galaxy can be gleaned from its luminosity profile (that is, its surface intensity as a function of radius). But in order to study the host-galaxy profiles

we must first remove the contribution from the quasar point source. To achieve this we need to estimate the quasar luminosity precisely enough to subtract its point-spread function (PSF) from the raw image to give an accurate image of the host galaxy. The main problems in achieving subtraction are: saturation of the central pixels, uncertainty in the local PSF, scattered light and photon noise. There are several ways to proceed, none of them obviously *a priori* better than the rest. We have tried several, compared the results and carried out numerous simulations as checks. The procedures finally adopted have been driven by the decisions: (1) to get the most reliable data for the hosts rather than for the quasars, (2) to ignore the diffraction spikes which vary too much across the field of view for accurate subtraction; (3) to otherwise use as much of the unsaturated and uncontaminated data as possible.

To study the host profiles, we want to subtract a scaled PSF from the raw data. We have first attempted this by subtracting scaled versions of a model PSF from the raw data until a point is reached beyond which the residual 'galaxy image' ceases to increase monotonically inwards towards the most central unsaturated pixels. This method is intended to enable us to study the underlying galaxy in a rough but model-independent way. Figure 1 shows the luminosity profiles of the residual host images, both as a plot of $\log(\text{intensity})$ versus semimajor axis and as $\log(\text{intensity})$ versus $(\text{semimajor axis})^{1/4}$. Spiral galaxies would yield a straight line in the first plot, ellipticals in the second. It is clear from Fig. 1 that three of the quasars are well fitted by an elliptical profile but not a spiral whereas PKS1302-102 is more equivocal. Nonetheless, it is still better fitted by an elliptical than a spiral profile. This suggests the four hosts are ellipticals.

Assuming that a host galaxy has a luminosity profile which can be well fitted by a simple model (for example, $r^{1/4}$ (where r is the radius from the centre of the galaxy) or exponential) then more accurate parameters for both quasar and host can be obtained using the two-component cross-correlation method, details of which we have published elsewhere^{17,23}. This technique, which is entirely independent of the first, finds the best fit between a model PSF and a series of model galaxy templates. Using this method, we found in all four cases that $r^{1/4}$ models give much better fits to the data (in χ^2 terms) than exponential (spiral) profiles. We defined an 'acceptable fit' to mean that the value of χ^2 was such that the model could not be rejected at the 95% confidence level when tested against the actual image. In all four cases the best-fitting spiral model could be rejected at the 95% confidence level whereas the best-fitting elliptical model could not. This confirms the earlier findings of our model-independent subtraction technique (Fig. 1). The derived quasar/host parameters are listed in Table 1. We have tried enough alternatives and compared them with simulations to be confident that this procedure will yield results with systematic errors of $\leq 20\%$ in all quoted parameters (in Table 1) for our particular sample.

In the cosmology adopted here, an L^* galaxy (that is, a galaxy with a luminosity on the knee of the galaxy luminosity function) will have $M_V \approx -22$ and a brightest cluster member will have $M_V \approx -24.5$. Our host galaxies have M_V values of -22.8 to -23.7 (that is, they are 2-5 times brighter than L^* galaxies). This high luminosity has already been suggested by Veron-Cetty and Woltjer¹² for 13 radio-loud quasars studied from the ground.

Figure 2 presents grey-scale maps of the four host galaxies after the best-fitting quasars are removed. The scaling factors used to subtract the PSF from the raw images were those derived from the best fit using the cross-correlation method. Two different intensity scales are shown: the first illustrates the full extent of the host galaxies; the inset is a shallower cut designed to show the companion objects. Brief comments on each object follow.

PHL1093 is an optically selected object which has a steep-spectrum, lobe-dominated radio source²⁵. Ground-based studies^{7,9,11,15} have produced conflicting conclusions about both

* Present address: Department of Physics, University of Bristol, H. H. Wills Physics Laboratory, Tyndal Avenue, Bristol BS8 1TL, UK.

the galaxy size and its morphology. The host in our data seems to be a smooth elliptical fitting an $r^{1/4}$ profile exactly (Fig. 1a) and has a major axis of position angle $PA \approx 66^\circ$. In comparison, the radio lobes²⁵ have $PA \approx 25^\circ$. Our derived $L_{\text{nuc}}/L_{\text{gal}} = 0.62$ (where L_{nuc} is the luminosity of the nucleus, and L_{gal} is the luminosity of the galaxy) compares with ground-based values of 0.3 in B (ref. 9) and 0.3 in R (ref. 11). The HST data (Fig. 2a) show the presence of a previously undetected close (1.3 arcsec, ≈ 11 kpc projected separation) companion entirely within the host isophotes. The companion object has an apparent luminosity in the F702W filter band of $m_{\text{F702W}} = 21.1$. If it is at the quasar distance it has an absolute luminosity in the F702W filter band of $M_{\text{F702W}} = -20.1$.

Q2215-037 is a radio-quiet object that was first detected in X-rays. Of the many ground-based observations^{4, 6, 8-10, 15} some suggest that the host galaxy has 'elliptical characteristics'¹⁰ whereas others suggest spiral or interacting characteristics⁴. The HST again shows a host galaxy with elliptical morphology. Our value of $L_{\text{nuc}}/L_{\text{host}} = 0.73$ is much lower than those derived by any ground-based study, which range from 2.5 in R (ref. 10) to

1.6 in B (ref. 9) for an elliptical fit. Study of the Wide Field Camera frames, which accompany each PC frame, reveals that Q2215-037 is probably part of a poor cluster stretching 3 arcmin to the northwest and itself lies ~ 13.2 arcsec (112 kpc) away from an even brighter elliptical ($M_V = -23.0$ if lying at the same distance). The HST data (Fig. 2b) show a small, previously undetected, companion 1.6 arcsec from the centre of the host galaxy (a projected separation of 12.6 kpc) which is indistinguishable from a star, although the probability of finding a star of this brightness projected so close to the galaxy is small ($< 2\%$). It has $m_{\text{F702W}} = 23.7$. If at the distance of the quasar it must be a very faint dwarf ($M_{\text{F702W}} = -17.4$) or the tidally stripped remains of a once-larger companion.

PKS2128-123 is a higher-redshift radio-loud object included to test HST's ability to resolve hosts at redshift $z \approx 0.5$. It has a compact unresolved (with 4.5 arcsec resolution) radio structure with a flat spectrum²⁷. Our subtraction reveals a luminous host with $M_{\text{F702W}} = -23.7$ and effective radius $R_e = 37.4$ kpc, the large size enabling us to detect it at this redshift despite the presence of a quasar 13 times brighter than the host. The HST data (Fig.

FIG. 1 Luminosity profiles of the residual host galaxies left after subtraction of a scaled model PSF from the raw images. For each galaxy, the luminosity profile is shown as $\log(\text{counts})$ against semimajor axis (left) and also as $\log(\text{counts})$ against $(\text{semimajor axis})^{1/4}$ (right). An exponential profile would have a straight line in the left-hand plots while an $r^{1/4}$ -law profile would have a straight line in the right-hand plots. The profiles have been plotted out to the radius at which the signal-to-noise ratio in an isophote falls to 10. Each model PSF used was generated by TINY TIM software³². This was aligned to have its peak within 0.2 pixels of the peak in the raw image. Scaled versions of the PSF, of increasing intensity, were then subtracted, as described, from the raw image. Pixels contaminated by saturation, diffraction spikes or an obvious companion object were ignored in this process. Isophotal ellipses were then fitted to the resulting residual galaxy images using the ISOPHOTE package in the STSDAS software (see text). Pixels contaminated by saturation, diffraction spikes or an obvious companion object were again ignored in this process. The error bars represent ± 1 standard deviation of the mean intensity within each isophotal ellipse.

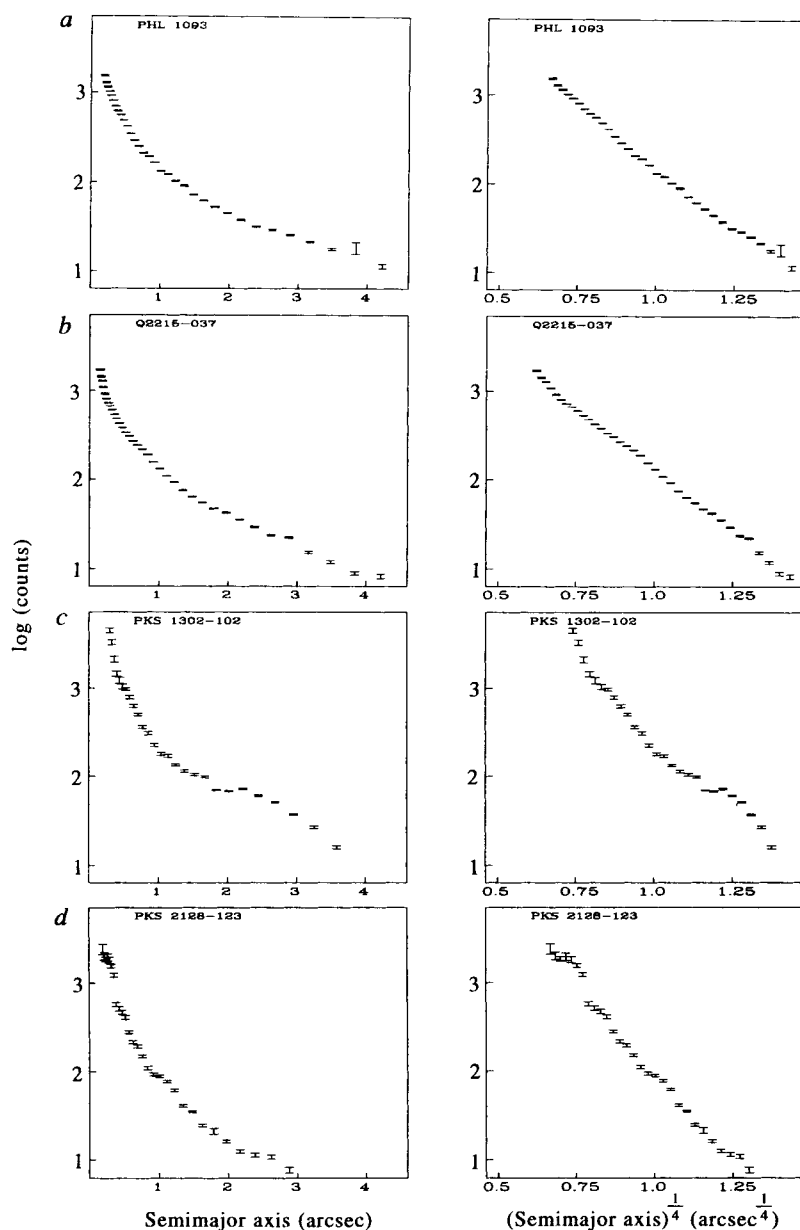


TABLE 1 Parameters of the four host galaxies

Object	Class	z	V	M_V	m_{quas}	m_{host}	a/b	PA	M_{quas}	M_{host}	R_e (kpc)	$\frac{L_{\text{nuc}}}{L_{\text{host}}}$
PHL1093	Radio-loud	0.258	17.07	-23.9	18.3	17.8	0.88	66°	-22.9	-23.4	24.0	0.62
Q2215-037	Radio-quiet	0.241	17.20	-23.7	18.6	18.2	0.80	118°	-22.5	-22.8	5.4	0.73
PKS1302-102	Radio-loud	0.286	15.20	-26.0	16.0	18.0	0.87	22°	-25.5	-23.5	10.5	6.26
PKS2128-123	Radio-loud	0.501	16.11	-26.6	16.3	19.1	0.91	139°	-26.6	-23.7	37.4	13.54

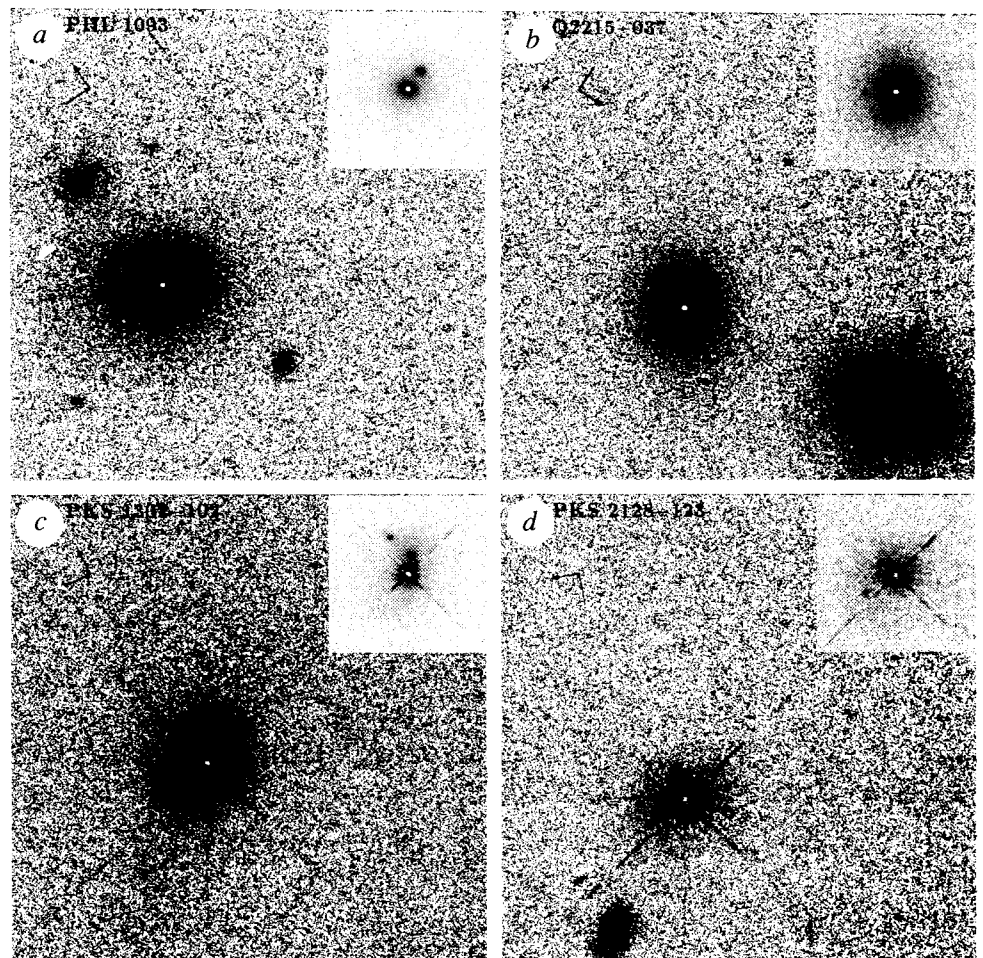
Columns 1–5 list the properties of the quasars taken from the ground-based measurements of Veron-Cetty and Veron²¹. The M_V values were K-corrected (that is, corrected for redshifting a different part of the intrinsic spectrum into our filter) by Veron-Cetty and Veron as described in ref. 21. Columns 6–13 list the properties of the quasars and host galaxies as determined by the present study. All magnitudes refer to the F702W filter. Parameters listed here were derived using the two-component cross-correlation technique^{17,27}. This technique finds the best fit between a model PSF template and a series of galaxy model templates. It selects not only that galaxy model which gives the best χ^2 fit but also the relative scaling factors of the PSF model and galaxy model which give this best fit. In each case $r^{1/4}$ (elliptical) profiles were found to give a better fit (in χ^2 terms) than exponential (spiral) ones. Parameters in this table refer to the best-fitting $r^{1/4}$ profile. Regions of the image contaminated by saturated pixels, diffraction spikes or an obvious companion object were ignored in this process. Questionable aspects of the cross-correlation procedure include: (1) force-fitting host profiles to idealized $r^{1/4}$ or exponential profile; (2) the use of theoretical as opposed to real PSFs (to minimize problems of variable diffraction spikes, shot-noise and spacecraft ‘breathing’); (3) ignoring scattered light. (Based on an STSCI report³¹ on scattered light, we calculate that the only one of our objects in which it will be significant is PKS2128-123, the object with the largest $L_{\text{gal}}/L_{\text{nuc}}$. In PKS2128-123 it may approach ~25% of the intensity of the detected host galaxy outside radii >2 arcsec. At $r=0.5$ arcsec it contributes only 15%. However, ignoring pixels outside $r=2$ arcsec has a tiny effect on the derived galaxy parameters.) The M_{F702W} values were obtained using standard HST calibration files with no K-correction applied. All but one of the quasars has a redshift such that its rest-frame V emission is redshifted into the F702W filter. A comparison of total F702W magnitude (quasar + host) derived from our model fitting with the ground-based K-corrected M_V values shows that the two sets match reasonably well ($M_{\text{F702W}} \approx M_V + 0.3 (\pm 0.2)$). Hence we feel safe in comparing our values of M_{F702W} with V-band parameters of nearby galaxies. a/b and PA were measured from the host galaxy images presented in Fig. 2. a/b was measured at the effective radius R_e . Many simulations suggest that the combined errors should be $<20\%$ for all quoted figures.

2d) reveal a close companion (1.9 arcsec, 34.6 kpc projected separation). The companion has $m_{\text{F702W}} = 23.9$ ($M_{\text{F702W}} = -19.4$) and is clearly extended. Several absorption-line systems are seen in the spectrum of this quasar and one at $z=0.42$ has been associated with a bright spiral galaxy 8.6 arcsec away^{24,28}. If our

close companion is part of the same cluster as the spiral it could be the absorber instead. A redshift is needed to test this possibility.

PKS1302-102 is a radio-loud object studied extensively from the ground^{2,4,5,11,12,14} and with the aberrated WFPC²⁶. The radio

FIG. 2 Grey-scale plots of the four quasars after subtracting a model PSF from the raw images. Two different intensity scales have been chosen for each object: one (main picture) to best show the host galaxies and the other (inset) to best show any companion objects. Each of the large panels is 27.6×27.6 arcsec. The model PSF subtracted here was identical to that used in the determination of the profiles shown in Fig. 1 except that the scaling factor used here was determined using the two-component cross-correlation technique^{17,23} described in Table 1. Note that the diffraction spikes remain. The outermost clearly detectable isophote of a host galaxy typically has a surface brightness of 24.3 mag arcsec⁻² (in the F702W magnitude system).



source has a small two-sided structure (<10 arcsec) resembling a miniature Wide-Angled Tail source²⁶. We find the host (Fig. 1c) is better fitted by an $r^{1/4}$ (elliptical) than an exponential profile although the errors are manifestly larger than in the other three host galaxies. The derived value of $L_{\text{nuc}}/L_{\text{host}} = 6.3$ is slightly higher than values of $L_{\text{nuc}}/L_{\text{host}}$ obtained from data taken using an R filter, as found in ground-based work (4.7 and 2.5; refs 4 and 11, respectively). The host galaxy has $\text{PA} \approx 22^\circ$. This is similar to the position angle of the northern radio lobe of PKS1302–102 (ref. 26) ($\text{PA} \approx 17^\circ$). The host (Fig. 1c) has two close companions, one at 1.1 arcsec (projected separation, 10 kpc), the other at 2.3 arcsec (projected separation, 21 kpc). The latter has been detected from the ground¹⁴ and the former by the aberrated WFPC²⁶; both have also been detected in recent WFC observations²⁰. The two companion objects have $m_{\text{F702W}} = 21.3$ ($M_{\text{F702W}} = -20.2$) and $m_{\text{F702W}} = 22.0$ ($M_{\text{F702W}} = -19.4$) respectively.

PKS1302–102 is the one object we have in common with Bahcall *et al.*²⁰ who are observing a similar sample with the lower-resolution HST WFC. Their data also show the two companions we report here. They do not, however, show the host galaxy. From simulations they calculate that they would only detect a large elliptical host galaxy under PKS1302–102 if it had $m_{\text{F606W}} < 18.3$. The host galaxy we find has $m_{\text{F702W}} = 18.0$. If it has $V-R$ colour typical of elliptical galaxies then it will have $m_{\text{F606W}} \approx 18.5$ and hence would not have been detected by Bahcall *et al.*

Our observations demonstrate the capability of the HST to contribute significantly to our understanding of quasar environments and possibly quasar physics. We cannot draw statistical conclusions from this tiny sample. Nevertheless our four quasars (three radio-loud, one radio-quiet) seem to have elliptical rather than spiral hosts as does the one unequivocal quasar of Hutchings *et al.*^{29,30}. They are 2–5 times brighter than L^* galaxies but not as luminous as brightest cluster members. It will be interesting to see if these trends continue. All four quasars show circumstantial evidence for interaction, in that all have at least one very close companion – too close to be chance alignments in general. Combined with the finding of Bahcall *et al.*²⁰ (five out of eight with companions), this lends credence to the long-standing idea that the quasar phenomenon is related to the fuelling of the central source in the nucleus of a host galaxy by matter falling onto it during an interaction or merger. $\square$

27. Morganti, R., Killeen, N. E. B. & Tadhunter, C. M. *Mon. Not. R. astr. Soc.* **263**, 1023–1048 (1993).
28. Boissé, P., Boulade, O., Kunth, D., Tytler, D. & Vigroux, I. *Astr. Astrophys.* **262**, 401–416 (1992).
29. Hutchings, J. B. *et al. Astrophys. J.* **429**, L1–L4 (1994).
30. Hutchings, J. B. & Morris, S. C. *Astr. J.* **109**, 1541–1545 (1995).
31. Krist, J. & Burrows, C. *Large Angle Scattering in WFPC2 and Horizontal 'Smearing' Correction* (WFPC2 Instrument Science Report 94-01, STSCI, Baltimore, 1994).
32. Krist, J. *Astronomical Data Analysis Software and Systems II* (eds Hanish, R. J., Brissenden, R. J. V. & Barnes, J.) (ASP Conf. Ser. 52, San Francisco, 1993).

ACKNOWLEDGEMENTS. We thank J. Trauger and his team for producing a fine camera. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA Inc., under contract to NASA. F.M. was on assignment from the Space Science Department of ESA. This work was supported by the European Space Agency.

Cenozoic variations in the flux of interplanetary dust recorded by ^3He in a deep-sea sediment

K. A. Farley

Division of Geological and Planetary Sciences, MS 170-25,
California Institute of Technology, Pasadena, California 91125, USA

HELIUM-3 concentrations and $^3\text{He}/^4\text{He}$ ratios in modern pelagic sediments are known to be far in excess of terrestrial values as a result of micrometeorite fallout^{1–3}. Here I report that extraterrestrial helium is easily detected in a pelagic clay core dating back more than 70 Myr. The remarkable preservation of the extraterrestrial signature arises from high retention of ^3He within interplanetary dust particles, coupled with loss of radiogenic ^4He from terrestrial mineral grains. The core provides a continuous record of the fallout of extraterrestrial helium for the Cenozoic era. This record suggests that there have been significant variations in the influx of interplanetary dust through time, probably related to asteroidal breakup events or the passage of comets through the inner Solar System. The results also show ^3He to be a far more sensitive tracer of the interplanetary dust flux than is iridium.

Core LL-44-GPC-3 is from the pelagic clay province of the central North Pacific and is thought to carry a nearly complete sedimentation record to ~ 72 Myr ago⁴. The sediment is dominated by aeolian and hydrothermal components which accumulated at slow but variable rates in response to climate conditions and the migration of the site across the Pacific basin^{4,5}.

The ratio $^3\text{He}/^4\text{He}$ in the GPC-3 sediments is in the range $(3.2\text{--}207) \times 10^{-6}$ (Table 1, Fig. 1) compared with values of $\sim 200 \times 10^{-6}$ in pure interplanetary dust⁶. The extraterrestrial fraction of ^3He in these sediments can be calculated from the measured $^3\text{He}/^4\text{He}$ ratio and the known composition of other helium sources likely to be present: crustal helium ($^3\text{He}/^4\text{He} < 0.02 \times 10^{-6}$, ref. 7) atmosphere (1.4×10^{-6}) and mantle helium ($< 42 \times 10^{-6}$, ref. 8). Given the high He concentration and the general dearth of volcanic material in the core⁵, the presence of mantle helium trapped in volcanic glass can be ignored. Similarly, the He/Ne ratio in these sediments was always more than 100 times the atmospheric value, precluding the existence of significant amounts of atmospheric helium in the samples. Sediment helium is simply a binary mixture of crustal and extraterrestrial components, as proposed previously⁹. Using reasonable endmember $^3\text{He}/^4\text{He}$ ratios⁹, the fraction of ^3He which is extraterrestrial always exceeds 98%, demonstrating that ^3He is extraordinarily sensitive to extraterrestrial input. The survival of such high $^3\text{He}/^4\text{He}$ ratios and ^3He abundances in samples up to 70 Myr old is surprising given the high concentration of ^4He -producing uranium and thorium in these sediments⁴, and the supposedly high diffusivity of ^3He from extraterrestrial matter on the sea floor^{10,11}.

Received 28 February; accepted 2 June 1995.

1. Kristian, J. *Astrophys. J.* **179**, L61–L64 (1973).
2. Wyckoff, S., Wehinger, P. A. & Gehren, T. *Astrophys. J.* **247**, 750–761 (1981).
3. Hutchings, J. B. & Campbell, B. *Nature* **303**, 584–588 (1983).
4. Hutchings, J. B., Crampton, D. & Campbell, B. *Astrophys. J.* **280**, 41–50 (1984).
5. Hutchings, J. B., Crampton, D., Campbell, B., Duncan, D. & Glendenning, B. *Astrophys. J. Suppl. Ser.* **55**, 319–366 (1984).
6. Malkan, M. A., Margon, B. & Chanan, G. A. *Astrophys. J.* **280**, 66–78 (1984).
7. Malkan, M. A. *Astrophys. J.* **287**, 555–565 (1984).
8. Margon, B., Downes, R. A. & Chanan, G. A. *Astrophys. J. Suppl. Ser.* **59**, 23–31 (1985).
9. Smith, E. P., Heckman, T. M., Bothun, G. D., Romanishin, W. R. & Balick, B. *Astrophys. J.* **306**, 64–89 (1986).
10. Hutchings, J. B., Janson, T. & Neff, S. G. *Astrophys. J.* **342**, 660–665 (1989).
11. Hutchings, J. B. & Neff, S. G. *Astr. J.* **99**, 1715–1721 (1990).
12. Veron-Cetty, M.-P. & Wolter, L. *Astr. Astrophys.* **236**, 69–85 (1990).
13. Neff, S. G. & Hutchings, J. B. *Astr. J.* **100**, 1441–1451 (1990).
14. Hutchings, J. B. & Neff, S. G. *Astr. J.* **104**, 1–14 (1992).
15. Dunlop, J. S., Taylor, G. L., Hughes, D. H. & Robson, E. I. *Mon. Not. R. astr. Soc.* **264**, 455–488 (1993).
16. Abrahams, R. G., Crawford, C. S. & McHardy, I. M. *Astrophys. J.* **401**, 474–480 (1992).
17. Boyce, P. J., Philipps, S. & Davies, J. I. *Astr. Astrophys.* **280**, 694–703 (1993).
18. Miller, J. S. in *Astrophysics of Active Galaxies and Quasi-Stellar Objects* (ed. Miller, J. S.) (Oxford Univ. Press, Oxford, 1985).
19. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. *Astrophys. J.* **435**, L11–L14 (1994).
20. Bahcall, J. N., Kirhakos, S. & Schneider, D. P. *Astrophys. J.* (in the press).
21. Veron-Cetty, M. P. & Veron, P. A. *Catalogue of Quasars and Active Nuclei* 6th edn (Sci. Rep. No. 13, ESO, Garching, 1993).
22. Burrows, C. J. *Hubble Space Telescope Wide Field and Planetary Camera 2 Instrument Handbook, Version 2.0* (STSCI, Baltimore, 1994).
23. Philipps, S. & Boyce, P. J. *Mon. Not. R. astr. Soc.* **256**, 673–678 (1992).
24. Bergeron, J. *Astr. Astrophys.* **155**, L8–L11 (1986).
25. Gower, A. C. & Hutchings, J. B. *Publ. astr. Soc. Pacif.* **96**, 19–23 (1984).
26. Hutchings, J. B., Morris, S. C., Gower, A. C. & Lister, M. L. *Publ. astr. Soc. Pacif.* **106**, 642–645 (1994).

The concentration of ^4He decreases systematically between the surface and the Cretaceous/Tertiary boundary (66 Myr), then rises slightly at the bottom of the core (Fig. 1a). The measured concentrations are up to two orders of magnitude lower than predicted from closed-system ingrowth in the sediment column (Fig. 1), conclusively demonstrating ^4He escape¹², probably by diffusion out of submicrometre-sized⁵ mineral grains. As a consequence, the high $^3\text{He}/^4\text{He}$ extraterrestrial signature remains largely undiluted by ^4He ingrowth for at least 70 Myr. The observed down-hole trend in $[\text{He}]$, including the increase in the deepest two samples, correlates with the fraction of aeolian sediment⁴, indicating that ^4He is a tracer of aged (^4He -rich) continental detritus.

In contrast, $[\text{He}]$ rises steadily down-hole to a sharp maximum at 11.4 m (~37 Myr), drops dramatically to a minimum at 12 m (~39 Myr), rises again to a second maximum at ~15 m (~50 Myr), then decreases steadily to the bottom of the hole (Fig. 1). The maximum concentration of $125 \times 10^{-12} \text{ cm}^3 \text{ STP g}^{-1}$ at 11.4 m depth is higher by a factor of 3 than sediments analysed previously^{3,9} and is associated with an unusually high abundance of cosmic spherules (F. T. Kyte, personal communication). These spherules are thought to be melted micrometeorites, but are probably not the carrier of the extraterrestrial helium. Because infalling particles larger than ~50 μm diameter suffer frictional heating and outgassing during atmospheric entry, the extraterrestrial helium is predominantly carried within small unmelted interplanetary dust particles³ (IDPs).

Laboratory diffusion measurements extrapolated to sea-floor temperatures predict complete diffusive loss of ^3He from IDP magnetite grains in <10⁷ years after fallout^{10,11}. The GPC-3 data demonstrate that the natural diffusivity from IDPs must be much lower, probably by several orders of magnitude. This discrepancy may reflect extrapolation from a laboratory-determined diffusivity which is invalid under sea-floor conditions, or, alternatively, may signal the existence of a more He-retentive phase than magnetite within the IDP population. The existence of an extraterrestrial silicate phase, hosting ~50% of the ^3He , has been postulated previously¹⁵, but its helium retentivity is unknown.

In addition to relatively slow diffusivity, the data dictate that a large fraction of the IDP complex is resistant to sea-floor chemical weathering.

The concentration of extraterrestrial ^3He in a sediment is governed by the relation:

$$[\text{He}] = \frac{fR}{\alpha}$$

where f is the flux of IDP ^3He , R is the fractional retention of IDP ^3He after arrival on the sea bed (which probably decreases monotonically with age) and α is the sediment mass accumulation rate. Essentially all ^3He is extraterrestrial in the GPC-3 samples, so the observed $[\text{He}]$ profile carries information on both the IDP helium influx and on α , provided that the various effects can be deconvolved.

The accumulation history of GPC-3 has been estimated by both biostratigraphic and geochemical means. The former is likely to be more accurate, but provides a limited number of time horizons, whereas the latter is more uncertain, but provides a continuous record. Using these approaches for approximating α , it is possible to assess the implied ^3He flux (that is, the product fR), and to evaluate qualitatively the flux and retentivity individually. Magnetostratigraphy provides ages in the uppermost 2.5 Myr of GPC-3, but the magnetic signal is lost below this depth¹⁴. For the remainder of the core, epoch boundaries are identified by ichthyolith assemblages and by the 'spike' of iridium concentration that initiates the Tertiary period⁴ (Fig. 2). Using the total sediment accumulation⁴ and ^3He abundance for each epoch, a coarse-resolution history of the implied ^3He flux can be calculated (Fig. 2). In the six Cenozoic epochs, this value varies in the range $(1.2-0.34) \times 10^{-12} \text{ cm}^3 \text{ STP cm}^{-2} \text{ kyr}^{-1}$, with a mean of $0.6 \times 10^{-12} \text{ cm}^3 \text{ STP cm}^{-2} \text{ kyr}^{-1}$. The highest implied ^3He flux, in the Pleistocene, is similar to previous estimates for the Quaternary⁹. Throughout the Tertiary, the implied flux (in $10^{-12} \text{ cm}^3 \text{ STP cm}^{-2} \text{ kyr}^{-1}$) is much lower, averaging 0.48 and varying between 0.34 in the Miocene and 0.63 in the Oligocene epochs.

The data in Fig. 2b preclude large, long-lived fluctuations in the flux of IDP helium to the Earth in the Cenozoic era, and clearly confirm that ^3He retention must remain high over a 10⁸-year timescale. The dramatically higher flux in the Quaternary relative to the Tertiary is subject to two interpretations. It could reflect real variations in the flux of IDP ^3He , or could be a consequence of an initially rapid reduction in retentivity, followed by near-constant retention. Because there are multiple carrier phases of extraterrestrial ^3He in sediments^{10,13}, a step-function in retentivity could arise from rapid loss of ^3He from one phase coupled with high retention in a second. However, experiments in progress (K.A.F., unpublished results) do not support a change in ^3He carrier between Quaternary and Tertiary samples. In any case the relative constancy of the epoch-averaged implied flux through the Tertiary suggests that R remains nearly constant over this period, as a fortuitous balance of flux and retentivity variations would otherwise be required.

If retentivity effects are secondary, then observed inter-epoch variations in implied flux must track changes in the infall of extraterrestrial helium to the Earth. Averaged over epoch intervals, the variations in the Tertiary are less than a factor of two in total, but the entire span occurs between the Miocene and the Oligocene. Because the implied flux increases with age between these epochs, this trend cannot result from retentivity effects. It seems necessary to conclude that the ^3He flux in the Tertiary reached a maximum during the Oligocene, then decreased by almost a factor of two into the Miocene.

Kyte and co-workers⁴ have developed a detailed sedimentation history for GPC-3 by combining magnetostratigraphy with the assumption that hydrogenous cobalt accumulates on the sea floor at a constant rate. This 'constant-Co' model matches the biostratigraphic data reasonably well (Fig. 2a) and allows the

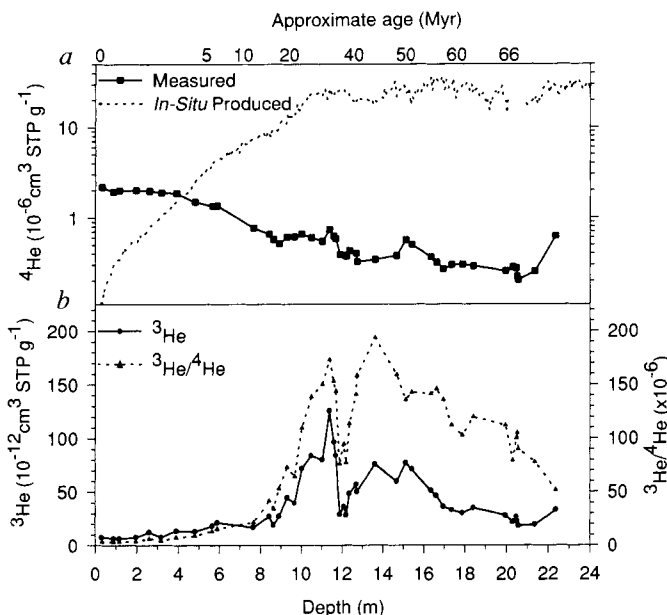


FIG. 1 a, Measured $[\text{He}]$ in core GPC-3 compared with expected *in situ* production given the age and U and Th abundance⁴ of the sediments. b, $[\text{He}]$ and $^3\text{He}/^4\text{He}$ ratios in GPC-3. In both plots the analytical uncertainty is smaller than the symbols.

TABLE 1 Helium-isotope data from core LL-44-GPC-3

Interval (cm)	Age (Myr)	^3He ($10^{-12} \text{ cm}^3 \text{ STP g}^{-1}$)	$^3\text{He}/^4\text{He}$ ($\times 10^{-6}$)	Interval (cm)	Age (Myr)	^3He ($10^{-12} \text{ cm}^3 \text{ STP g}^{-1}$)	$^3\text{He}/^4\text{He}$ ($\times 10^{-6}$)
32–36	0.19	7.7	3.5	1,220–1,224	39.04	28.4	75.9
87–96	0.43	6.5	3.4	1,236–1,240	39.70	47.3	112.0
115–123	0.54	6.3	3.2	Replicate		49.5	114.9
119–200	0.85	7.4	3.7	1,263–1,271	40.89	57.0	141.1
260–270	1.16	12.3	6.2	1,273–1,276	41.27	50.1	154.2
317–321	1.50	7.6	4.0	1,360–1,365	45.18	103.4	206.9
385–394	2.08	13.7	7.4	Replicate-1		68.8	189.8
476–485	3.37	12.9	8.7	Replicate-2		71.1	194.1
565–570	5.85	18.1	13.5	Replicate-3		60.0	184.3
590–600	6.66	21.4	15.8	1,466–1,470	49.04	59.8	157.6
767–771	12.63	16.9	21.9	1,506–1,514	50.61	77.2	136.0
843–848	15.97	27.4	41.7	1,540–1,544	51.93	70.9	141.1
864–868	17.01	19.1	33.2	1,634–1,640	55.76	51.2	140.2
891–897	18.65	27.1	52.5	1,661–1,664	56.70	46.4	144.8
928–935	21.10	44.7	73.0	1,695–1,698	57.83	36.4	135.4
968–971	23.94	39.6	64.5	1,735–1,740	59.29	33.0	110.5
1,003–1,007	26.70	71.8	109.0	1,787–1,791	61.49	30.2	99.7
1,050–1,054	30.41	83.6	139.7	1,840–1,846	63.36	35.1	120.9
1,098–1,108	33.18	79.9	145.5	1,995–1,998	65.80	28.1	109.9
1,138–1,143	35.14	129.7	176.3	2,030–2,033	66.15	21.8	77.5
Replicate		121.4	164.0	2,046–2,049	66.32	26.9	98.1
1,159–1,164	36.26	96.7	155.5	2,051–2,054	66.36	22.7	102.9
1,165–1,172	36.69	84.0	144.0	2,056–2,059	66.40	18.3	89.4
1,188–1,191	37.59	29.2	75.0	2,136–2,140	67.40	19.3	76.1
1,203–1,211	38.46	36.5	95.1	2,235–2,240	69.27	33.1	52.1

The core is from $30^\circ 19.9' \text{ N}$, $157^\circ 49.9' \text{ W}$, 5,705 m water depth. Age assignments are from Kyte *et al.*⁴. He analyses were made on $\sim 50 \text{ mg}$ of bulk sediment after fusion at $1,200^\circ \text{C}$, using conventional mass spectrometric techniques at Caltech. Blank levels of $\sim 5 \times 10^{-11} \text{ cm}^3 \text{ STP}$ of ^4He and $2 \times 10^{-15} \text{ cm}^3 \text{ STP}$ of ^3He were insignificant ($<2\%$) in all cases. Replicates generally agreed to within $\pm 5\%$ (with the notable exception of one analysis at 1,360 cm). This variability is far larger than analytical uncertainty ($<2\%$), and probably indicates sediment heterogeneity, as previously observed⁹.

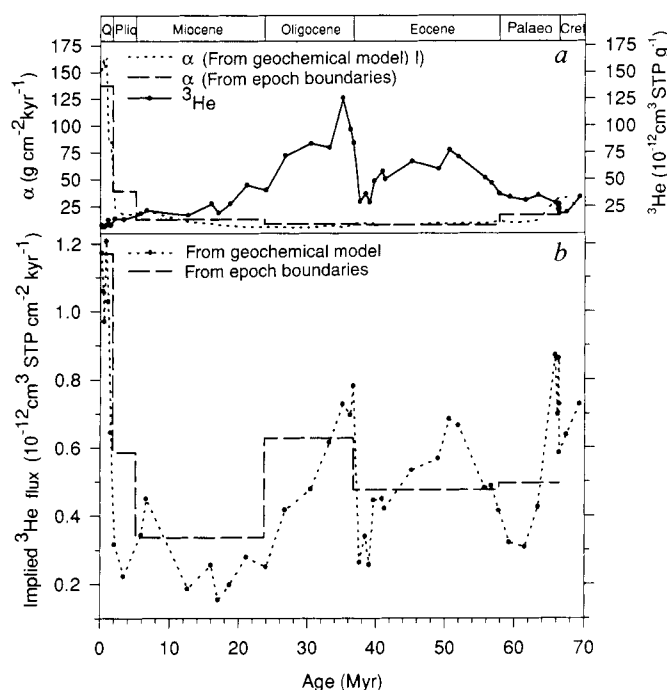


FIG. 2 a, Sediment mass accumulation rates (α) determined from epoch boundaries and by using the 'constant-Co' geochemical model⁴, with ^3He data for comparison. b, Implied ^3He flux (that is, the product of the flux and retention factor), calculated using the two mass accumulation rate profiles in a. The average implied flux calculated from the geochemical model within a given epoch may differ slightly from the value calculated using the epoch boundaries as a result of small discrepancies in the assumed position of the boundaries in the two models for α (ref. 4).

calculation of a continuous record of the implied ^3He flux for the Cenozoic. The continuous profile shows significant short-term variability which is lost when epoch averages are calculated. For example, the minimum value (at 18 Myr) is a factor of 6 lower than the highest (at $\sim 1 \text{ Myr}$), and the implied flux oscillates over a factor of about 5 with a period of the order of 20 Myr. Note that these fluctuations cannot be produced by variations in R (which should decrease monotonically with age), and, within the framework of the sedimentation rate model, can only imply variations in the flux of IDP ^3He .

The Cretaceous/Tertiary (K/T) boundary at GPC-3 is easily identified by a large Ir peak⁴, but is associated with essentially no $[\text{He}]$ peak (Fig. 1b). There is a small (40%) spike in the implied ^3He flux at the boundary, but this is far less than the 40-fold increase observed⁴ in Ir. Furthermore, the helium spike arises mostly from the sedimentation model, which is least reliable at precisely this time as a result of the injection of impact-derived detritus⁴. Thus even the small increase in implied flux coincident with the Ir peak may be an artefact. Away from the K/T boundary, the Ir profile in GPC-3 shows no obvious structure, just a broad, weak maximum from 8 to 18 m depth which correlates with sedimentation rate⁴. The concentration of ^3He is generally elevated over this same interval, but shows a great deal more structure.

Decoupling of ^3He and Ir in the GPC-3 sediments illustrates the fact that the two tracers of extraterrestrial fallout are delivered to the sea floor from different sources and in different ways. Most importantly, ^3He can record only direct fallout of dust grains because larger bodies (for example, the K/T impactor) experience impact devolatilization. This size bias is further enhanced by the fact that He may be implanted by energetic ions in space^{10,15,16}, so initial He concentrations are probably highest in small bodies with high surface/volume ratios. In contrast Ir records direct and indirect fallout of bodies of any size, and may be chemically mobilized and reprecipitated. The relative sensitivities of ^3He and Ir to the IDP flux can be calculated from an estimate of the IDP composition and flux. Taking the ^3He

concentration observed in individual stratospheric IDPs ($1.7 \times 10^{-5} \text{ cm}^3 \text{ STP g}^{-1}$, ref. 6), the most ^3He -rich GPC-3 sediment (1,138–1,143 cm) carries $\sim 7 \text{ p.p.m.}$ by weight of IDPs small enough to retain helium. The He-bearing fraction corresponds to $\sim 1\%$ of the total IDP flux¹⁷, so the total IDP component in this interval is $\sim 700 \text{ p.p.m.}$ Assuming a chondritic Ir concentration of 481 p.p.m. (ref. 18), this 700 p.p.m. of IDPs contributes just 0.3 p.p.b. of Ir to the bulk sediment. A concentration of 0.3 p.p.b. amounts to only $\sim 15\%$ of the total Ir observed in the interval ($\sim 2 \text{ p.p.b.}$, ref. 4) and is within the uncertainty of the Ir measurements⁴. This calculation demonstrates that ^3He is much more sensitive to the influx of extraterrestrial dust than is Ir.

Interplanetary dust reaching the Earth's atmosphere is thought to derive from both asteroidal and cometary debris in the Zodiacal Cloud^{19,6}. Although it is likely that gravitational effects favour both greater capture¹⁹ and lower degrees of heating²⁰ and He outgassing⁶ of asteroidal particles, both sources are thought to contribute to the present-day helium flux⁶. The observed ^3He flux variations are most simply attributed to changes in the abundance of dust in the Zodiacal Cloud associated with asteroidal breakup events and the passage of comets through the inner Solar System. In addition, changes in the solar output of energetic ^3He ions implanted into IDPs may modulate the average ^3He concentration in the infalling dust.

One of the most dramatic features in the implied flux record occurs near the end of the Eocene epoch. Between 37.6 and 36.3 Myr the implied flux nearly triples, a rise which is essentially indistinguishable from a step-function given bioturbation effects in the core. Within the age uncertainty of samples, this rise is synchronous with the occurrence of several tektite horizons observed around the world²¹. This correlation provides strong evidence for multiple terrestrial impact events associated with dust-bearing objects, possibly a comet shower. Evidence has also been presented for an unusually high abundance of impact craters in the Quaternary²², another period in which the implied ^3He flux is high. Further study of well-dated Cenozoic sediments is required to determine the origin of the ^3He flux variations and to evaluate the relationship between the ^3He record and known terrestrial impact events²². In particular a more detailed record of the rise and fall rates of the ^3He flux, and their temporal relation to evidence for impacts, would be very useful.

It has previously been suggested that subduction of IDP-bearing deep-sea sediments may account for the ^3He emitted by volcanoes²³, but this has been disputed both because helium can diffuse from IDPs during subduction¹¹ and because the absolute abundance of extraterrestrial helium in sediment seems insufficient^{24,25}. Although the GPC-3 data demonstrate that high $^3\text{He}/^4\text{He}$ ratios persist in deep-sea sediments for many millions of years, the flux of ^3He is actually lower than previously assumed²³. The GPC-3 results demonstrate that, throughout the Cenozoic era, the flux of IDP ^3He has been insufficient to supply the current emission from the mantle²⁴ even if all of the helium were subducted, and provides no new support for the idea that extraterrestrial helium plays a role in mantle geochemistry. □

14. Prince, R. A., Heath, G. R. & Kominz, M. *Geol. Soc. Am. Bull.* **91**, 1789–1835 (1980).
15. Tilles, D. *Science* **153**, 981–985 (1966).
16. Nier, A. O., Schlutter, D. J. & Brownlee, D. E. *Geochim. cosmochim. Acta* **54**, 173–182 (1990).
17. Love, S. G. & Brownlee, D. E. *Science* **262**, 550–553 (1993).
18. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).
19. Flynn, G. J. *Planet. Space Sci.* **42**, 1151–1161 (1994).
20. Flynn, G. J. *Icarus* **77**, 287–310 (1989).
21. Keller, G., D'Hondt, S. & Vallier, T. *Science* **221**, 150–152 (1983).
22. Hut, P. et al. *Nature* **329**, 119–127 (1987).
23. Anderson, D. L. *Science* **261**, 170–174 (1993).
24. Trull, T. in *Noble Gas Geochemistry and Cosmochemistry* (ed. Matsuda, J.) 77–88 (Terrapub, Tokyo, 1994).
25. Stuart, F. M. *Earth planet. Sci. Lett.* **122**, 245–247 (1994).

ACKNOWLEDGEMENTS. I thank F. Kyte, K. Turekian and B. J. Pegrum for helpful comments and for supplying samples, and G. J. Flynn for a helpful review. This work was supported by the Ocean Sciences Division of the US National Science Foundation.

Unusual twentieth-century summer warmth in a 1,000-year temperature record from Siberia

Keith R. Briffa*, Philip D. Jones*, Fritz H. Schweingruber†, Stepan G. Shiyatov‡ & Edward R. Cook§

* Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

† Swiss Federal Institute of Forest, Snow and Landscape Research, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

‡ Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, 8 Marta Street, Ekaterinburg 620219, Russia

§ Tree-Ring Laboratory, Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

In the current debate on the magnitude of modern-day climate change, there is a growing appreciation of the importance of long, high-resolution proxies of past climate^{1–3}. Such records provide an indication of natural (pre-anthropogenic) climate variability, either singly at specific geographical locations or in combination on continental and perhaps even hemispheric scales⁴. There are, however, relatively few records that are well dated, of high resolution and of verifiable fidelity in terms of climate response, and conspicuously few that extend over a thousand years or more⁵. Here we report a tree-ring-based reconstruction of mean summer temperatures over the northern Urals since AD 914. This record shows that the mean temperature of the twentieth century (1901–90) is higher than during any similar period since AD 914.

Ring-width measurements from living and subfossil siberian larch (*Larix sibirica*) trees have previously been used to produce a continuous series representing interannual growth variability in a region centred on about 66° 50' N, 65° 15' E on the eastern side of the northern Urals, near Salekhard⁶. This mean ring-width chronology, running from AD 961 to AD 1969, formed the basis of a high-summer (June–July) season temperature reconstruction⁷. However, this reconstruction is of limited value in understanding 100-year timescale fluctuations, because of the way individual tree measurements were detrended to remove age-related sample bias in the mean chronology^{8,9} (a process known as 'standardization' in dendroclimatology^{9,10}). We have now resampled these trees, and analysed both living and subfossil material using wood densitometry to yield both ring-width and maximum-latewood-density time series¹¹ spanning the years from AD 914 to AD 1990. The density data were standardized so as to preserve variance at both short (interannual-to-decadal) and long (>100 years) timescales¹². The ring-width data were standardized using flexible smoothing splines, producing

Received 15 March; accepted 9 June 1995.

1. Merrihue, C. *Ann. N.Y. Acad. Sci.* **119**, 351–367 (1964).
2. Kyrlov, A. Y., Manyrin, B. A., Silin, Y. I. & Khabarin, L. V. *Geochim. Int.* **10**, 202–205 (1973).
3. Ozima, M., Takayanagi, M., Zashu, S. & Amari, S. *Nature* **311**, 449–451 (1984).
4. Kyte, F. T., Leinen, M., Heath, G. R. & Zhou, L. *Geochim. cosmochim. Acta* **57**, 1719–1740 (1993).
5. Corliss, B. H. & Hollister, C. D. in *The Ocean Floor* (eds Scrutton, R. A. & Talwani, M.) 277–304 (Wiley, New York, 1982).
6. Nier, A. O. & Schlutter, D. J. *Meteorites* **25**, 263–267 (1990).
7. Andrews, J. N. *Chem. Geol.* **49**, 339–351 (1985).
8. Kurz, M. D., Jenkins, W. J., Hart, S. R. & Clague, D. *Earth planet. Sci. Lett.* **66**, 388–399 (1983).
9. Takayanagi, M. & Ozima, M. *J. geophys. Res.* **92**, 12531–12538 (1987).
10. Amari, S. & Ozima, M. *Nature* **317**, 520–522 (1985).
11. Hiyagon, H. *Science* **263**, 1257–1259 (1994).
12. Podosek, F. A., Honda, M. & Ozima, M. *Geochim. cosmochim. Acta* **44**, 1875–1884 (1980).
13. Fukumoto, H., Nagao, K. & Matsuda, J. I. *Geochim. cosmochim. Acta* **50**, 2245–2253 (1986).

TABLE 1 Relationships between tree growth and regional temperatures in the northern Urals

(a) Tree-growth temperature correlations

	Previous year				Year of tree growth												A1	A2
	S	O	N	D	J	F	M	A	M	J	J	A	S	O				
Density																		
1882–1935	0.08	0.01	0.17	0.02	0.11	0.07	0.05	0.15	0.44†	0.68†	0.59†	0.27*	0.32*	0.02	0.15	0.20		
1936–1990	−0.14	0.23	0.01	0.08	0.12	0.37†	−0.04	0.17	0.35*	0.64†	0.67†	0.39†	0.36†	−0.07	0.15	0.27*		
Ring width																		
1882–1935	0.24	0.07	0.11	0.29*	0.22	−0.14	−0.06	0.08	0.02	0.43†	0.55†	−0.10	0.19	0.14	0.29*	0.40†		
1936–1990	−0.03	0.25	0.11	−0.10	0.13	0.35	−0.10	−0.07	0.18	0.38†	0.49†	0.17	0.09	0.03	0.45†	0.26		

(b) May–September mean temperature regressions

Calibration period	1882–1935	1936–1990	1882–1990
Verification period	1936–1990	1882–1935	—
Calibration			
Variance explained	0.70	0.66	0.68
Verification			
Variance explained	0.65	0.69	
Reduction of error	0.67	0.70	
Coefficient of efficiency	0.65	0.68	
First difference sign			
Correct	46	44	—
Incorrect	8	9	—
Product mean <i>t</i>	5.4	3.9	—
Regression weights			
Max. density year <i>t</i>	1.10	1.10	1.11
year <i>t</i> + 1	–0.11	–0.14	–0.14
Ring width year <i>t</i>	–0.48	–0.56	–0.54
year <i>t</i> + 1	0.20	0.34	0.27

The density chronology was constructed by averaging measured values for all individual trees expressed as anomalies from a common empirically defined linear expression of declining density as a function of age. This straight-line reference was calculated as a least-squares fit to the observed average densities of all samples after aligning them with respect to their cambial age, regardless of their calendrical year of growth. The resulting indices, subsequently averaged across correct calendrical years, provide a chronology in which long-timescale variability is preserved and sample-age bias is largely eliminated¹². Separate standardizing equations were calculated for the modern and subfossil density data to explore the possibility of a systematic difference in the density/age relationship that might arise due to a slight elevational difference in these data sources (the subfossil samples being up to 80 m higher than the living trees). Such a difference might introduce a bias between the early and recent sections of the chronology. The similarity of the density/age equations derived from the separate subfossil (914–1744) and modern (1631–1990) samples argue against any long-term sample-related bias in the density chronology, although this does not preclude the possibility that changes in the average altitude of the samples through time might introduce some shorter-timescale bias where the replication is relatively low²⁸. The width chronology is the average of series for each tree after low-frequency variance, associated with a smoothing spline fit to the original measurements, has been removed. The frequency responses of the splines are tailored to the length of each series so that most of the variance on timescales longer than two-thirds the series length is removed. This produces a final ring-width chronology with low error variance, but with a spectrum that is truncated at low frequencies (that is, periods >50–100 years)²⁹. Individual monthly mean temperature correlations are shown with each of the maximum-latewood-density and ring-width chronologies. Also shown are the results of multiple regressions in which both the density and ring-width chronologies are used as predictors of mean May–September temperature. The results of verification tests^{10,30} are all highly statistically significant. A1 and A2 are the 1-year and 2-year lag autocorrelation coefficients.

* Significant at the 5% level.

† Significant at the 1% level or better.

a chronology characterized by variance on shorter timescales (generally below 100 years⁹) (see Table 1).

These chronologies were then used in a number of regression experiments designed to exploit the well-known dependence of tree growth on summer temperature variability at high latitudes^{13–16}. This is demonstrated in Table 1, which shows the density and ring-width chronologies correlated with monthly mean temperatures averaged within a region between 62.5° and 67.5° N, and 65° to 75° E¹⁷. All significant correlations are positive, indicating enhanced width and density associated with warmer temperatures. Correlations with ring width are greatest in June and July. Density correlations are stronger and manifest over a longer (May–September) season. The responses are also consistent in different periods. Mean May–September temperatures are very highly correlated ($r=0.95$ over 1881 to 1990) with cumulative growing degree days (>5 °C) in this region¹⁸. Regression results in which mean May–September temperatures are estimated as a function of density and width data in both the concurrent and following years (that is, *t* and *t* + 1 respectively), exhibit extremely strong predictive power (Table 1).

The explained variance over the calibration period (that is, 66–70%) and most importantly the levels explained in independent verification of the calculated regression equations (65–69%), and also the general levels of statistical significance associated with a number of stringent tests of model performance (see Table 1), all rank among the highest reported in the tree-ring literature. Regression based on the density predictors alone also explains a high degree of temperature variance (~60%) but consistently

less than that achieved using density and ring widths in combination.

The series (>1,000 years) of reconstructed summer temperatures for the northern Urals shows extended periods of either relatively cool or relatively warm summers persisting over 100–200 years (Fig. 1). Superimposed on this is large variability on annual and decadal timescales (Table 2). Despite relatively cool summers in the late 1960s and 1970s, the mean summer temperature of the twentieth century (1901–90) appears to be the warmest of any 90-year period during the last ten centuries (0.13 °C warmer than the next warmest period, 1202–91). Most of the sixteenth and seventeenth centuries were notably cool, particularly the 1530s and 40s, the 1580s and 90s, and the 1620s, 30s and 40s. The pronounced coolness of the late sixteenth and seventeenth centuries and the warmth of the twentieth century is shown in other, though shorter, high-latitude summer-responsive proxies¹⁹ and in an average of data from 16 selected sites in the Northern Hemisphere⁴ (Fig. 2a). However, the coldest bidecadal mean in the northern Urals data, 1531–50, indicates an earlier onset of the cold period seen in the mean 'hemisphere' data and in a similar millennial-length temperature reconstruction for northern Fennoscandia based on tree-ring data from Torneträsk, northwest Sweden, that were also processed so as to preserve long-timescale climate variability¹².

Although summers were predominantly warm in northern Fennoscandia in the eleventh and twelfth centuries, they were clearly cool in the northern Urals (Fig. 2b). This is further support for the sparse body of high-quality palaeoclimate evidence

now challenging the still prevalent, over-simplified concept of a globally synchronous, multi-century Mediaeval Warm Period^{5,20}. But after the coolness of the late sixteenth century, both the northern Urals and the northern Fennoscandian temperature curves show overall warming. Linear trends calculated over 1600–1980 show a 0.67 °C warming in northern Fennoscandia, but 1.14 °C over the northern Urals. The two series also display qualitatively similar century-scale trends, warming to the mid-eighteenth century, slight cooling to the end of the nineteenth century, followed by an abrupt rise and sustained warmth during the twentieth century. The recent warmth in Fennoscandia is not significant in a 1,000-year context¹². The recent warmth in the Urals data is much larger (compare Table 2) but because of the century-timescale variability in the reconstruction, assessing its statistical significance is not straightforward. A comparison of the mean values for the pre- and post-1900 periods, using a *t*-test which takes account of lag-1 serial correlation²¹ in the sample data, indicates that the likelihood of the unusual warmth of the twentieth century arising by chance is less than 1 in 1,000. However, an empirically derived probability distribution based on multiple artificial series with the same persistence structure as the reconstruction (that is, AR3 rather than AR1) suggests a probability near 1 in 100. We also compared (using a simple *t*-test) the 1901–90 mean with that of the next warmest 90-year period (1202–91, compare Table 2). This stringent test indicates that the two periods are significantly different with a probability of 1 in 10. So although the twentieth century was certainly unusually warm in the northern Urals, determining how unusual it was, in the context of the long record, is equivocal.

These results and the above discussion may be relevant to general circulation model (GCM) studies of both 'natural' and anthropogenically perturbed climates. Recent coupled ocean–

atmosphere GCM experiments with anthropogenic forcing indicate that maximum warming is likely in high-latitude and continental regions^{22,23}. There is, of course, uncertainty in defining this expected anthropogenic pattern. At present, this is best viewed as the spatial temperature response of coupled ocean–atmosphere GCMs forced with gradually changing atmospheric compositions²⁴. Attempts to detect this response pattern in observational temperature data must take account of natural variability of climate on long (>50–100 years) timescales. Also, simulated variability from long GCM control runs must be checked against long palaeoclimate series such as the one we report here.

Ocean–atmosphere GCMs also suggest that areas bordering the North Atlantic will experience reduced warming and even localized cooling near regions of deep-water formation^{22,23}. A recent 600-year simulation of climate variability under present-day boundary conditions has revealed significant quasi-periodicity in annual temperatures centred on the northwest Atlantic with characteristic timescales of between 40 and 60 years (ref. 25). Similar variability, explicable in terms of oceanic thermohaline forcing, has been observed in a number of different regional observational records and has the potential to obscure anthropogenic warming in the region of the North Atlantic^{26,27}. The northern Fennoscandian summer temperatures show

TABLE 2 Extreme summer temperatures

(a) Individual summers

Warmest		Coldest	
Year	Anomaly (°C)	Year	Anomaly (°C)
1045	2.21	1032	−6.20
1033	2.11	1548	−4.73
1257	2.08	1538	−4.15
1471	1.96	1783	−4.08
1922	1.85	1816	−4.03
1464	1.84	1601	−4.01
1028	1.81	1531	−3.63
1468	1.70	1453	−3.62
1370	1.68	1523	−3.62
1720	1.65	1891	−3.55

(b) 20-year-mean periods

Warmest			Coldest		
Years	Anomaly	Standard error	Years	Anomaly	Standard error
1461–1480	0.45	0.27	1531–1550	−1.97	0.27
1948–1967	0.40	0.16	1623–1642	−1.85	0.26
1238–1257	0.34	0.17	998–1017	−1.75	0.24
1919–1938	0.24	0.23	1580–1599	−1.61	0.17
1355–1374	0.09	0.23	966–985	−1.44	0.29

(c) 50-year-mean periods

Warmest			Coldest		
Years	Anomaly	Standard error	Years	Anomaly	Standard error
1919–1968	0.23	0.14	1595–1644	−1.64	0.15
1208–1257	0.03	0.12	1522–1571	−1.55	0.15
1353–1402	−0.04	0.12	966–1015	−1.39	0.17
1460–1509	−0.13	0.14	1410–1459	−1.10	0.12
1261–1310	−0.18	0.13	1808–1857	−0.96	0.15

(d) 90-year-mean periods

Warmest			Coldest		
Years	Anomaly	Standard error	Years	Anomaly	Standard error
1901–1990	0.03	0.10	1522–1611	−1.52	0.10
1202–1291	−0.10	0.11	961–1050	−1.20	0.15
1301–1390	−0.28	0.10	1612–1701	−1.12	0.11

Extreme individual summer temperature estimates along with the extreme mean values from among all 20-year, 50-year and 90-year periods. Temperatures are shown as anomalies with respect to the 1951–70 period. Only values for non-overlapping periods are shown. Of the extreme individual summers, nine are reconstructed as cooler, and four warmer, than the reconstructed extremes (1891 and 1922) during the calibration period. Compared with the instrumental extremes (1891 and 1915), 15 summers are reconstructed as cooler but none warmer.

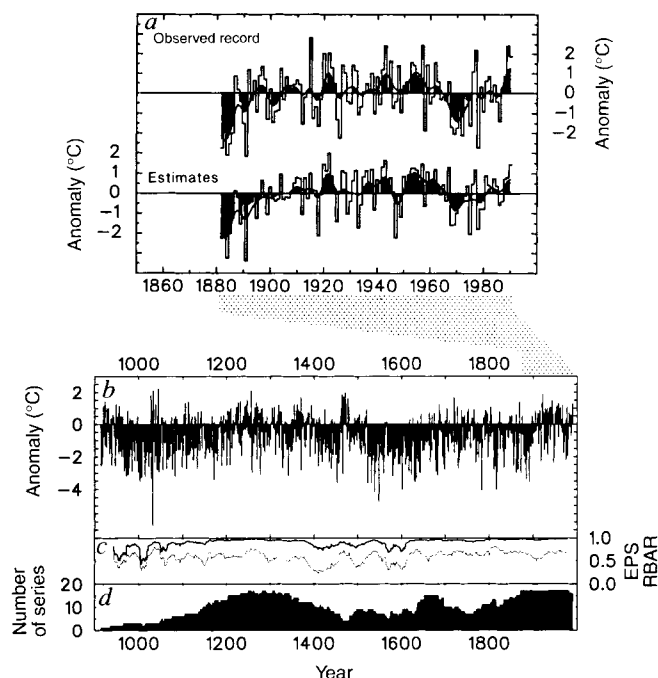


FIG. 1 *a*, Observed and estimated mean summer (May–September) temperatures over the northern Urals shown as anomalies (°C) from the 1951–70 mean. The smoothed line shows decadal-filtered values. *b*, The extended reconstruction (914–1990) expressed as anomalies from the same base. *c*, A measure of the strength of common growth 'signal' within the density chronology, the mean interseries correlation³¹ (RBAR; thin line) calculated between all samples over a moving 30-year window. The Expressed Population Signal (EPS; thick line) which is a function of RBAR and the series replication shown in *d*, indicates chronology reliability. Values >0.85 are generally considered satisfactory³¹.

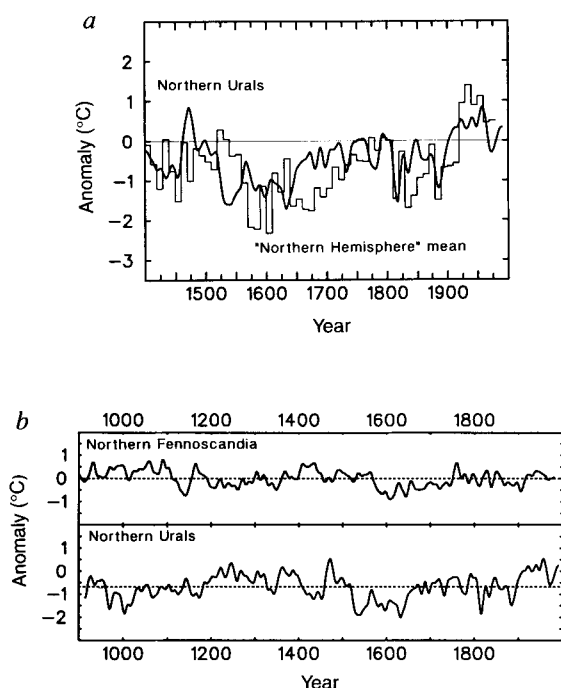


FIG. 2 *a*, The histogram shows 'mean Northern Hemisphere' summer temperatures⁴ constructed as standardized decadal averages of 16 proxy series, compared with northern Urals temperatures smoothed to emphasise decadal variability. Both series are expressed as anomalies from the 1860–1959 mean. The data from ref. 4 have been adjusted to give a standard deviation of one over the same period. *b*, Summer temperatures in northern Fennoscandia¹² and the northern Urals smoothed with a 25-year low-pass filter. The vertical axes are anomalies from the respective means for 1951–70. The dotted lines show the means for 901–1981 (northern Fennoscandia) and 914–1990 (northern Urals).

marked multi-decadal variance at periods near 33 years and over a range from 60 to 70 years (as well as longer periods) which may be associated with north Atlantic oceanic forcing^{12,26}.

Prolonged twentieth-century summer warmth over the Urals, and the slower rate of warming in recent centuries in northern Fennoscandia, are therefore at least consistent with the pattern of anthropogenically induced warming predicted by the models. Although the relative warmth through much of the thirteenth and fourteenth centuries means that statistical significance of the twentieth-century Urals warmth is equivocal, our results indicate that in this region no other period of this length has been warmer, at least during the past 1,000 years. A very limited number of high-resolution records indicate unprecedented recent warmth in other high-latitude regions, but this warmth is either observed over a shorter recent period or within a shorter overall timeframe. □

12. Briffa, K. R. et al. *Clim. Dyn.* **7**, 111–119 (1992).
13. Jacoby, G. C. & Cook, E. R. *Arct. Alp. Res.* **13**, 409–418 (1981).
14. Briffa, K. R., Jones, P. D., Pilcher, J. R. & Hughes, M. K. *Arct. Alp. Res.* **20**, 385–394 (1988).
15. Briffa, K. R. et al. *Nature* **346**, 434–439 (1990).
16. Schweingruber, F. H., Briffa, K. R. & Nogler, P. *Int. J. Biomet.* **37**, 151–169 (1993).
17. Jones, P. D. et al. *J. Clim. appl. Met.* **25**, 161–179 (1986).
18. Jones, P. D. & Briffa, K. R. *Int. J. Climatol.* (in the press).
19. D'Arrigo, R. D. & Jacoby, G. C. Jr in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 296–311 (Routledge, London, 1992).
20. Folland, C. K., Karl, T. R. & Vinnikov, K. Ya. in *Climate Change: the IPCC Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 195–238 (WMO/UNEP, Cambridge Univ. Press, 1990).
21. Zwiers, F. W. & von Storch, H. *J. Clim.* **8**, 336–351 (1995).
22. Cubasch, U. et al. *Clim. Dyn.* **8**, 55–69 (1992).
23. Gates, W. L., Mitchell, J. F. B., Boer, G. J., Cubasch, U. & Meleshko, V. P. in *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Callander, B. A. & Varney, S. K.) 97–134 (WMO/UNEP, Cambridge Univ. Press, 1992).
24. Santer, B. D. et al. *Clim. Dyn.* (in the press).
25. Delworth, T., Manabe, S. & Stouffer, R. J. *J. Clim.* **6**, 1993–2011 (1993).
26. Stocker, T. F. *Nature* **367**, 221–222 (1994).
27. Schlesinger, M. E. & Ramankutty, N. *Nature* **367**, 723–726 (1994).
28. Briffa, K. R., Jones, P. D., Schweingruber, F. H., Karlén, W. & Shiyatov, S. G. in *Climate Fluctuations and Forcing Mechanisms of the Last 2,000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) (Springer, Berlin, in the press).
29. Cook, E. R. & Briffa, K. R. in *Methods of Dendrochronology* (eds Cook, E. R. & Kairiukstis, L. A.) 153–162 (Kluwer/IASA, Dordrecht, 1990).
30. Cook, E. R., Briffa, K. R. & Jones, P. D. *Int. J. Climatol.* **14**, 379–402 (1994).
31. Wigley, T. M. L., Briffa, K. R. & Jones, P. D. *J. Clim. appl. Met.* **23**, 201–213 (1984).

ACKNOWLEDGEMENTS. We thank E. Schär for densitometric analysis, and M. K. Hughes, T. M. L. Wigley and F. W. Zwiers for comments that helped in rewriting an earlier version of the manuscript. This work was supported by the University of East Anglia Innovation Fund and the EC Environment Research Programme (K.R.B.), the Swiss National Science Foundation (F.H.S.) and the US Dept of Energy (P.D.J.).

Longevity of sub-continental mantle lithosphere from osmium isotope systematics in orogenic peridotite massifs

L. Relsberg* & J.-P. Lorand†

* Centre de Recherches Pétrographiques et Géochimiques (CRPG/CNRS), BP 20, 54501 Vandœuvre-lès-Nancy, France

† Laboratoire de Mineralogie, URA CNRS 716, 61 Rue Buffon, Museum National d'Histoire Naturelle, 75005 Paris, France

ATTEMPTS to understand the formation and evolution of the sub-continental lithospheric mantle (SCLM) have been hampered by the absence of reliable time constraints, reflecting a lack of appropriate isotopic dating techniques. The most commonly used methods, involving strontium, neodymium and lead isotopes, yield ambiguous results in mantle rocks, and show no relationship with magmatic processes, as the low concentrations of these elements make them susceptible to later metasomatic disturbance. Osmium, by contrast, is much more abundant in the mantle than in the crust¹, so that peridotite Os isotope ratios are largely immune to recent metasomatic imprints. This provides a way to date the magmatic processes that determine mantle major-element compositions². We present here two examples of striking correlations between ¹⁸⁷Os/¹⁸⁸Os and Al₂O₃ concentration in orogenic peridotites, and argue that these can be used to date the differentiation of the SCLM. The old ages obtained agree with associated lower-crustal Nd model ages^{3–5}, and indicate that—in these post-Archaeon terrains as well as in Archæon cratons^{2,6,7}—SCLM can remain isolated from the convecting mantle for more than a billion years.

Orogenic peridotite massifs are slices of upper mantle that have been tectonically emplaced into the crust. They contain small proportions of mafic veins and layers, but are dominated by peridotites, which range in composition from clinopyroxene-rich lherzolites to clinopyroxene-poor harzburgites. The two such massifs discussed here are the Ronda Ultramafic Complex of southern Spain and the eastern Pyrenean peridotites. Ronda, a large contiguous ultramafic massif in the Betic-Rif Cordillera,

Received 6 December 1994; accepted 19 June 1995.

1. Eddy, J. A. *Global IGBP Change Report No. 19* (International Geosphere-Biosphere Programme, Stockholm, 1992).
2. Bradley, R. S. (ed.) *Global Changes of the Past* (University Corporation for Atmospheric Res./Office of Interdisciplinary Earth Studies, Boulder, 1991).
3. Bradley, R. S. & Jones, P. D. (eds) *Climate Since A.D. 1500* (Routledge, London, 1992).
4. Bradley, R. S. & Jones, P. D. *The Holocene* **3**, 367–376 (1993).
5. Hughes, M. K. & Diaz, H. F. *Clim. Change* **26**, 109–142 (1994).
6. Shiyatov, S. G. *Dendrochronology in the Upper Tree Line in the Urals* (Nauka, Moscow, 1986) (in Russian).
7. Graybill, D. A. & Shiyatov, S. G. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 393–414 (Routledge, London, 1992).
8. Shiyatov, S. G. & Mazepa, V. S. in *Methods of Dendrochronology* Vol. 1 (eds Kairiukstis, L. A., Bednars, Z. & Feliksik, E.) 87–96 (IASA/Polish Acad. Sci.—Systems Res. Inst., Warsaw, 1987).
9. Cook, E. R., Briffa, K. R., Shiyatov, S. G. & Mazepa, V. S. in *Methods of Dendrochronology* (eds Cook, E. R. & Kairiukstis, L. A.) 104–123 (Kluwer/IASA, Dordrecht, 1990).
10. Fritts, H. C. *Tree Rings and Climate* (Academic, London, 1976).
11. Schweingruber, F. H. *Tree Rings: Basics and Applications of Dendrochronology* (Reidel, Dordrecht, 1988).

was emplaced about 22 Myr ago^{8,9}. Isotope ratios of Sr, Nd and Pb in Ronda peridotites are extremely diverse and do not correlate with major-element composition^{10–12}. In contrast, Reisberg *et al.*¹³ showed that a strong correlation exists between Os isotopes and Mg number (Mg #, defined as $\text{Mg}^{2+}/(\text{Mg}^{2+} + \text{Fe}^{+2})$), an index of extent of melt extraction from peridotite residues. An even stronger correlation is observed when the same data are plotted against bulk-rock Al_2O_3 content (Fig. 1a), another melt extraction index.

The eastern Pyrenean ultramafic massifs, emplaced into the crust about 95 Myr ago¹⁴, differ from Ronda in that they are composed of numerous small, distinct bodies, spread out along the Pyrenees. Metasomatic phases are much more prevalent than in Ronda, and a few massifs (especially Caussou) have very high amphibole contents and clinopyroxene/orthopyroxene ratios >1 (ref. 15). Sr, Nd and Pb isotopes vary widely^{16,17}. Os isotopes (Table 1) are plotted against Al_2O_3 content in Fig. 1c. Despite the fact that these data include samples from six distinct eastern Pyrenean massifs, a strong correlation is observed among 16 of 17 samples analysed, including the highly metasomatized Caussou samples. The trend differs from that observed for Ronda in both slope and intercept.

The $^{187}\text{Os}/^{188}\text{Os}$ – Al_2O_3 correlations are most simply explained as isochron analogues, with Al_2O_3 concentration proportional to Re/Os. Because Re and Al are both moderately incompatible elements, whereas Os is compatible, melt extraction should, at least approximately, create correlations between Al and Re/Os in peridotite residues. With time, assuming initial isotope homogenization, radiogenic ingrowth of ^{187}Os from ^{187}Re creates linear relationships ("isochrons") between $^{187}\text{Os}/^{188}\text{Os}$ and $^{187}\text{Re}/^{188}\text{Os}$ if the system is unperturbed. However Re concentrations may be affected by recent metasomatism (see below). As Al is less easily perturbed, $^{187}\text{Os}/^{188}\text{Os}$ may be more strongly correlated with Al_2O_3 than with Re/Os. As in an isochron, the slope of the Os–Al trend reflects the time since isotope homogenization, while the y -intercept represents the initial isotope ratio. Such isotope homogenization may be caused by magmatism occurring when the ultramafic body left the convecting mantle and became incorporated in the SCLM and thereby sheltered

from convective disruption. The y -intercept thus represents the Os isotope ratio of the convective mantle at this time. In this model, isotope homogenization is required only over mantle volumes sufficiently large to approximate the Os mantle evolution curve. It is not required, for example, over the entire mass of the Ronda massif. The scatter observed in the trends may represent, in part, original mantle heterogeneities not completely erased by homogenization.

The time of lithospheric incorporation can be estimated by comparison of the y -intercept with the mantle evolution curve (Fig. 2). Although this curve is poorly constrained and may vary spatially¹⁸, most abyssal peridotites, ophiolites and recent ultramafic intrusions have $^{187}\text{Os}/^{188}\text{Os}$ ratios of 0.124–0.130 (refs 18–22). Assuming linear mantle evolution to this range from an initial ratio of 0.096–0.098 (ref. 23), Ronda is found to have a lithospheric age of 0.9–1.7 Gyr while the eastern Pyrenean massifs have an age of 1.9–2.6 Gyr. The correlations themselves may place tighter limits on the age estimates, assuming that the Os isotope ratios of fertile samples (~3.9% Al_2O_3) indicate the applicable present-day $^{187}\text{Os}/^{188}\text{Os}$ ratio for a given locality. For Ronda, 3.9% Al_2O_3 corresponds to $^{187}\text{Os}/^{188}\text{Os} \approx 0.127$; thus the y -intercept suggests an age of 1.3 Gyr (Fig. 2). For the eastern Pyrenean massifs, 3.9% Al_2O_3 corresponds to $^{187}\text{Os}/^{188}\text{Os} \approx 0.129$. This is slightly higher than most estimates of the upper-mantle Os composition, perhaps pointing to a contribution from recycled crust or the lower mantle, if the Os composition of the lower mantle is higher than that of the upper mantle²⁰. This ratio suggests an Os evolution curve yielding a preferred age of 2.3 Gyr. The two fertile samples (DES 7 and TUR 7) from the western and central Pyrenees plot slightly below the east Pyrenean trend, suggesting derivation from a more Re-depleted mantle segment. The sample plotting markedly off the correlation (FON-3) comes from Fontete Rouge, which unlike other eastern Pyrenean massifs, is protogranular, Ti-pargasite free and strongly light-rare-earth depleted^{16,24}. Thus it may reflect a later period of lithospheric addition.

The older age of the Pyrenees is consistent with the steeper slope of its Os–Al trend. Assuming a (poorly-constrained) $^{187}\text{Re}/^{188}\text{Os}$ ratio ~0.40 for fertile samples², the slopes suggest ages of

FIG. 1 a, Plot of $^{187}\text{Os}/^{188}\text{Os}$ ratio versus Al_2O_3 concentration (bulk rock) in Ronda peridotites. Os data from ref. 13. Filled squares denote massive peridotites, open symbols denote thin bands rich in garnet (diamond symbol) or spinel (triangle). Fitted line ($r^2 = 0.970$) excludes open triangle, for consistency with b, but visually indistinguishable lines are obtained if this sample is included, or if the fit includes only massive peridotites. Shaded field represents mixtures of harzburgite with 4 p.p.b. Os and plausible mantle liquids with $^{187}\text{Os}/^{188}\text{Os} = 0.15$ (lower bound, basalt with 0.01 p.p.b. Os, 14% Al_2O_3 , upper bound, picrite with 1 p.p.b. Os, 10% Al_2O_3). b, $^{187}\text{Os}/^{188}\text{Os}$ versus $^{187}\text{Re}/^{188}\text{Os}$ in Ronda peridotites. Symbols as in a. Fitted line ($r^2 = 0.834$) yields an 'age' of 1.24 Gyr. Open triangle excluded from fit for reasons given in text. c, $^{187}\text{Os}/^{188}\text{Os}$ versus Al_2O_3 (bulk rock) in peridotites from Pyrenean ultramafic massifs. Line ($r^2 = 0.948$) is a fit through filled squares, which denote East Pyrenean samples. Open squares represent samples from Fontête Rouge, an East Pyrenean massif with unusual characteristics. Open circles denote samples from central and western Pyrenees (TUR 7 and DES 7). Shaded field indicates plausible harzburgite–basalt mixtures (parameters as in a). d, $^{187}\text{Os}/^{188}\text{Os}$ versus $^{187}\text{Re}/^{188}\text{Os}$ in East Pyrenean peridotites. Symbols as in c.

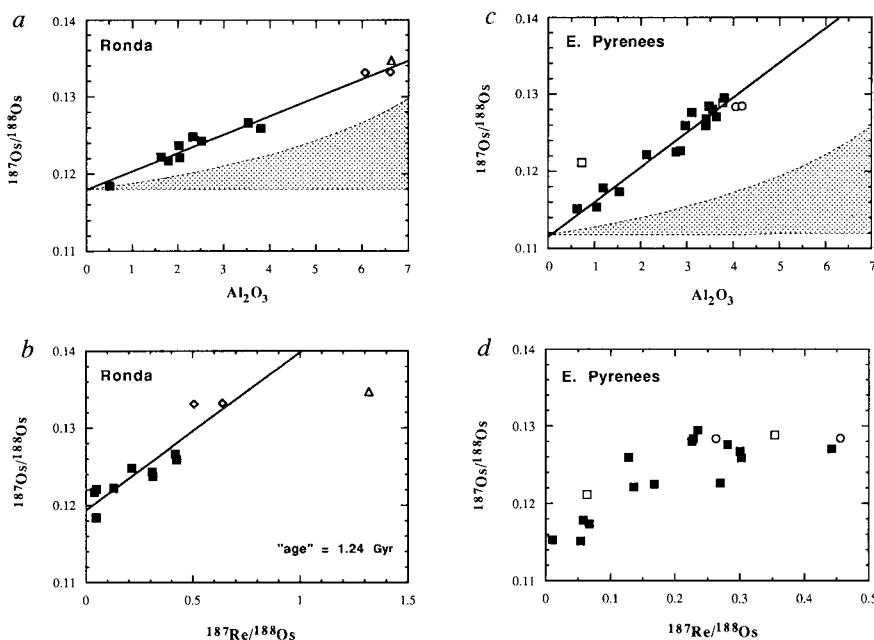


TABLE 1 Osmium isotope data from Pyrenean peridotites

Sample	Massif	Al ₂ O ₃ (wt %)	S (p.p.m.)	¹⁸⁷ Os/ ¹⁸⁸ Os	[Os] (p.p.b.)	[Re] (p.p.b.)	¹⁸⁷ Re/ ¹⁸⁸ Os
71-321	Lherz	3.10	220	0.1276	4.53	0.267 (0.328)	0.281 (0.345)
71-324	Lherz	2.96	170	0.1259	3.92	0.248 0.247	0.302 0.301
71-322	Lherz	1.53	56	0.1173	3.91	0.055	0.067
72-442	Lherz	1.18	20	0.1178	2.89	0.035	0.058
73-104	Lherz	1.04	100	0.1153 0.1142	5.95 3.69	0.013	0.010
71-325	Lherz	0.62	<20	0.1151	3.86	0.044	0.054
71-336	Freychinède	3.47	179	0.1284	3.43	0.164	0.228
71-339	Freychinède	3.41	244	0.1268	3.47	0.218	0.300
71-335	Freychinède	2.86	312	0.1226	4.12	0.233	0.270
72-425	Freychinède	2.77	223	0.1225	4.15	0.146 0.152	0.168 0.174
FON-4	Fontête Rouge	3.77	200	0.1288	3.81	0.283	0.354
FON-3	Fontête Rouge	0.72	46	0.1211 0.1226	2.77 2.11	0.037 (0.055)	0.064 (0.095)
70-354	Sem	3.80	170	0.1295	3.35	0.165	0.235
Sem 2	Sem	3.63	389	0.1271 0.1264	3.93 4.22	0.364	0.442
70-116	Caussou	3.55	300	0.128	3.52	0.167	0.226
70-5	Caussou	3.40	160	0.1259	3.05	0.082 (0.095)	0.128 (0.149)
71-264	Pic de Géal	2.13	100	0.1221	4.12	0.118 (0.129)	0.136 (0.149)
DES 7	Tuc Desse	4.05	280	0.1284	3.66	0.202	0.263
TUR 7	Turon de Téc.	4.18	270	0.1284	3.87	0.370	0.456

Massif locations given in ref. 24. Sulphur concentrations from ref. 36. Al₂O₃ contents determined by inductively coupled plasma (ICP) emission. Os samples digested by a standard NiS fusion technique (for example, ref. 10). Os was extracted by double distillation¹³, followed by purification on a single bead of Chelex 20 resin. Separate powder splits were dissolved in 3:1 HF-HNO₃ for Re analysis, then extracted using two columns of AG1X8 resin followed by a single bead of AG1X4 resin. Both Os and Re were analysed by negative thermal ionization mass spectrometry (NTIMS)^{37,38}. Precision on all ¹⁸⁷Os/¹⁸⁸Os measurements smaller than the reproducibility of standards (2σ = 0.3%). Total Os blank size was inversely correlated with blank isotope ratio, leading to a nearly constant correction of ~0.4% which has been applied to the listed isotope ratios. Re concentrations were corrected for a total blank of 12±8 pg per g of sample. Uncertainty on this blank correction provides the largest source of error in the listed Re concentrations. Repeat analyses were made on separate powder splits. Repeat Re concentrations in parentheses are from poor runs, and should be regarded as maximum values. The only one of six repeat Re analyses which does not agree within ±0.01 p.p.b. comes from a poor-quality run. Thus Re concentrations are quite reproducible, implying that use of separate powder splits for Re and Os analyses does not explain the lack of correlation in Fig. 1d.

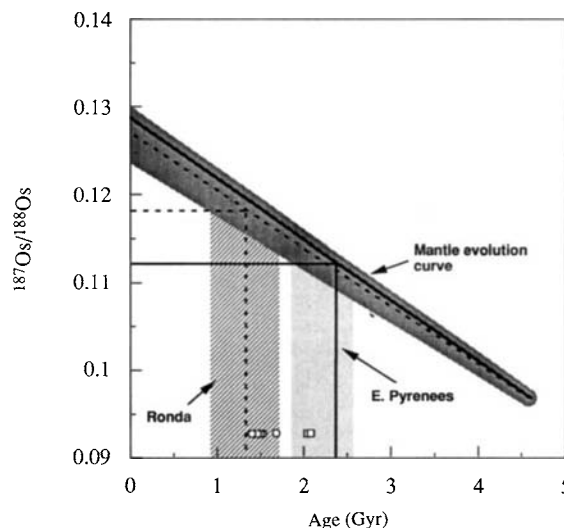
~1.36 Gyr (Ronda) and ~2.58 Gyr (Pyrenees), comparable to the independent estimates obtained from the y-intercepts. Other lines of evidence support the 1.3 Gyr age estimate for Ronda. All peridotites but one plot on a 1.24 Gyr Re/Os "errorchron" (Fig. 1b). (The excluded sample was very Cr-spinel rich and, unlike most others, digested under reducing conditions, suggesting underestimation of the Os concentration.) Further confirmation comes from the Re-Os errorchron obtained from mafic-layer samples (1.30 Gyr)¹³ and from Sm-Nd systematics of Ronda river sediment clinopyroxenes (1.3 Gyr)¹². Finally, in Beni Bousera (North Morocco), a closely related, contemporaneous Betic-Rif massif, Re-Os model ages from a thick pyroxenite layer converge to 1.2–1.3 Gyr (ref. 25) and the Sm-Nd model age of an exceptionally depleted peridotite lies²⁵ within error of 1.3 Gyr.

When the Pyrenean Re-Os data are plotted on an isochron diagram (Fig. 1d), the dispersion is too great to allow useful age information to be obtained, despite a rough positive correlation. The good reproducibility of Re concentrations and the limited range of most Os concentrations indicate that most of this scatter results neither from analytical difficulties nor from sample heterogeneity. The most likely explanation is that Re concentrations were perturbed by recent metasomatism. This view is supported by the poor ¹⁸⁷Os/¹⁸⁸Os correlation with sulphur content (Table 1), which strongly resembles the relationship with ¹⁸⁷Re/¹⁸⁸Os. In theory, the correlation with S should be better than that with Al₂O₃ because Re and Os are highly chalcophile. This suggests fairly recent S perturbation, perhaps associated with Ti-pargasitic metasomatism²⁶, which may also have perturbed Re and possibly Os concentrations.

Alternative origins for the Os-Al correlations must also be explored. One possibility might be mixing between basalt and harzburgite. This possibility can be rejected because the great difference in Os concentration between endmembers would produce strongly curved mixing trends, plotting well below the markedly linear observed trends (Fig. 1), even if extremely high (1 p.p.b.) magmatic Os concentrations are used. In any case, the unradiogenic isotopic composition required for the harzburgite endmember is most simply interpreted in terms of time of SCLM incorporation. A second alternative is that the trends result from continuous processes in the convecting mantle, analogous to those proposed by Allègre *et al.*²⁷ to explain apparent "mantle isochrons" in oceanic basalts. In this case the slopes and intercepts of the Os-Al correlations would have no simple time significance. However, the 1.3 Gyr age obtained from the y-intercept of the Ronda data is confirmed by several independent methods suggesting that it results from a well defined event. In addition, the trends defined by Ronda and the eastern Pyrenean massifs differ in slope and intercept, which would be surprising if both were recently extracted from the convecting mantle.

Thus the simplest explanation for the Al-Os correlations, radiogenic ingrowth since incorporation in the SCLM, is also the most plausible. Nevertheless, the marked linearity of the trends needs explanation. Simple melt-extraction models predict linearity only if Al and Re partition coefficients are nearly identical. The concentrations of both Re²⁸ and Al in MORBS are ~3 to 5 times higher than in fertile lherzolite, suggesting that their bulk partition coefficients are often comparable during mantle melting. But as Re is highly chalcophile whereas Al is lithophile, variations in sulphur systematics could cause the bulk partition

FIG. 2 Os mantle evolution curve. The y-axis intercepts of $^{187}\text{Os}/^{188}\text{Os}$ versus Al_2O_3 correlations are used to determine model ages for formation of mantle lithosphere beneath Ronda and the Pyrenees. Shaded regions indicate age ranges, taking into account reasonable variations in the mantle evolution curve. Dotted line (Ronda) and solid line (eastern Pyrenees) indicate best estimates of the appropriate mantle evolution curves to use in each case, based on the $^{187}\text{Os}/^{188}\text{Os}$ value at 3.9% Al_2O_3 . Open circles (Betic-Rif³) and squares (Pyrenees^{4,5}) indicate Nd model ages of associated lower-crustal samples.



coefficients of Re and Al to differ. Thus the observed linearity may not be explained solely by roughly similar melt/residue partitioning. Linear correlations are typically observed among major, compatible and moderately incompatible elements in orogenic peridotites²⁹. The reasons for this linearity, which may involve melt-lherzolite interaction and/or melt retention²⁹, are not yet well understood. Nevertheless, it is not too surprising that Re follows the pattern of most other elements, including sulphur²⁶. The Ronda samples with 6–7% Al_2O_3 , which are found in thin bands with distinct Sr and Nd compositions¹², probably reflect the passage of melts.

The data presented here suggest that the SCLM stabilized during a magmatic event, or series of events closely spaced in time, and thus argue against lithospheric formation by passive cooling of the asthenosphere. Rough agreement with Nd model ages obtained from associated lower crust suggests that SCLM formation and crustal extraction may have been linked (Fig. 2). In the Betic-Rif orogen, four out of five metapelite gneisses surrounding the Beni Bousera massif yield³ Nd model ages of 1.40–1.56 Gyr. Considering the uncertainties inherent in model age calculations, these ages are consistent with the lithospheric age suggested by the Os-Al relationship in Ronda. In the

Pyrenees, zircon U-Pb concordia upper-intercept ages^{30,31} and Nd model ages of Hercynian silicic igneous rocks^{4,5,32} fall mostly between 1.4 and 2.1 Gyr. Because these samples may include mantle components added at the time of emplacement, these ages represent minimum estimates for the lower crust. Better estimates may come from two Pyrenean lower-crustal xenoliths, which yield^{4,5} Nd model ages of 2.0–2.1 Gyr in good agreement with the time of lithospheric formation estimated from the Pyrenean peridotites.

The old ages derived from the Os-Al trends imply long-term survival of the SCLM. In Archaean root zones, dominated by depleted garnet harzburgites with low Fe contents, stability is attributed to density contrasts with the convecting mantle³³. As orogenic peridotites are dominated by spinel lherzolites, which are normally more Fe-rich³⁴, density contrasts relative to the asthenosphere should be much smaller. Both Ronda and the Pyrenees are in tectonically active regions. In Ronda, Mesozoic compressional lithospheric thickening was followed by mid-Tertiary uplift³⁵. In the Pyrenees, evidence exists for repeated episodes of deformation and magmatic injection²⁴. Despite this tectono-magmatic disruption, in both regions the SCLM survived for over a billion years. □

Received 25 May; accepted 22 June 1995.

- Morgan, J. W. & Lovering, J. F. *Earth planet. Sci. Lett.* **3**, 219–224 (1967).
- Walker, R. J., Carlson, R. W., Shirey, S. B. & Boyd, F. R. *Geochim. cosmochim. Acta* **53**, 1583–1595 (1989).
- Polvé, M. thesis, Univ. Paris VII (1983).
- Bickle, M. J., Wickham, S. M., Chapman, H. J. & Taylor, H. P. *Contr. Miner. Petrol.* **100**, 399–417 (1988).
- Gilbert, J. S., Bickle, M. J. & Chapman, H. J. *Chem. Geol.* **111**, 207–226 (1994).
- Carlson, R. W. & Irving, A. J. *Earth planet. Sci. Lett.* **126**, 457–472 (1994).
- Pearson, D. G. et al. *Geochim. cosmochim. Acta* **59**, 959–977 (1995).
- Priem, H. N. A. et al. *Contr. Miner. Petrol.* **70**, 103–109 (1979).
- Zindler, A., Staudigel, H., Hart, S. R., Endres, R. & Goldstein, S. *Nature* **304**, 226–230 (1983).
- Hauri, E. thesis, WHOI/MIT (1992).
- Reisberg, L. & Zindler, A. *Earth planet. Sci. Lett.* **81**, 29–45 (1987).
- Reisberg, L., Zindler, A. & Jagoutz, E. *Earth planet. Sci. Lett.* **96**, 161–180 (1989).
- Reisberg, L. C., Allègre, C.-J. & Luck, J.-M. *Earth planet. Sci. Lett.* **105**, 196–213 (1991).
- Albarède, F. & Michard-Vitrac, A. *Earth planet. Sci. Lett.* **40**, 327–332 (1978).
- Fabries, J., Bodinier, J.-L., Dupuy, C., Lorand, J.-P. & Benkerrou, C. *J. Petrology* **30**, 199–228 (1989).
- Downes, H., Bodinier, J.-L., Thirlwall, M. F., Lorand, J.-P. & Fabries, J. *J. Petrology* (Spec. vol. orogenic lherzolites and mantle processes) 97–115 (1991).
- Mukasa, S. B., Shervais, J. W., Wilshire, H. G. & Nielson, J. E. *J. Petrology* (Spec. vol. orogenic lherzolites and mantle processes) 117–134 (1991).
- Hattori, K. & Hart, S. R. *Earth planet. Sci. Lett.* **107**, 499–514 (1991).
- Roy-Barman, M. & Allègre, C. J. *Geochim. cosmochim. Acta* **58**, 5043–5054 (1994).
- Martin, C. E. *Geochim. cosmochim. Acta* **55**, 1421–1434 (1991).
- Snow, J. E. & Reisberg, L. *Earth planet. Sci. Lett.* (in the press).
- Luck, J.-M. & Allègre, C. J. *Earth planet. Sci. Lett.* **107**, 406–415 (1991).

- Luck, J.-M. & Allègre, C. J. *Nature* **302**, 130–132 (1983).
- Fabries, J., Lorand, J.-P., Bodinier, J.-L. & Dupuy, C. *J. Petrology* (Spec. vol. orogenic lherzolites and mantle processes) 55–76 (1991).
- Kumar, N., Reisberg, L. & Zindler, A. *Geochim. cosmochim. Acta* (submitted).
- Lorand, J.-P. *J. Petrology* (Spec. vol. orogenic lherzolites and mantle processes) 77–95 (1991).
- Allègre, C. J., Brévar, O., Dupré, B. & Minster, J. *Phil. Trans. R. Soc. A* **297**, 447–477 (1980).
- Morgan, J. W. *J. geophys. Res.* **91**, 12375–12387 (1986).
- Frey, F. A., Suen, C. J. & Stockman, H. W. *Geochim. cosmochim. Acta* **49**, 2469–2491 (1985).
- Briqueu, L. & Innocent, C. *C.r. hebd. Seanc. Acad. Sci. Paris* **316**, 623–628 (1993).
- Delaperrière, E., de Saint Blanquat, M., Brunel, M. & Lancelot, J. *Bull. Soc. géol. Fr.* **165**, 101–112 (1994).
- Ben Othman, D., Fourcade, S. & Allègre, C. J. *Earth planet. Sci. Lett.* **69**, 290–300 (1984).
- Jordan, T. H. in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics* (eds Boyd, F. R. & Meyer, H. O. A.) 1–14 (Am. Geophys. Union, Washington DC, 1979).
- McDonough, W. F. *Earth planet. Sci. Lett.* **101**, 1–8 (1990).
- Van der Wal, D. thesis, Univ. Utrecht (1993).
- Lorand, J.-P. *Earth planet. Sci. Lett.* **93**, 50–64 (1989).
- Creaser, R. A., Papanastassiou, D. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **55**, 397–401 (1991).
- Volkening, J., Walczyk, T. & Heumann, K. G. *Int. J. Mass Spec. Ion Phys.* **105**, 147–159 (1991).

ACKNOWLEDGEMENTS. We thank C. Allègre and J.-M. Luck for their insight and collaboration during the study of the Ronda Massif, and A. Zindler and J. Fabries for many fruitful discussions. D. Dole and L. Pattou provided technical assistance. This Letter benefited from reviews by H. Becker, G. Pearson, M. Chaussidon, J. Ludden and J. Snow. The work was supported by a CNRS grant (Action Spécifique 'Science de l'Univers').

Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements

Jeffrey L. Boore*, Timothy M. Collins,
David Stanton*, L. Lynne Daehler
& Wesley M. Brown

Department of Biology, University of Michigan,
830 N. University Avenue, Ann Arbor, Michigan 48109-1048, USA

THE origins of arthropods and the phylogenetic relationships among their three major living groups (atelocerates, crustaceans and chelicerates) are vigorously contended. To help resolve this, we determined mitochondrial gene arrangements for a chelicerate, a myriapod, two crustaceans, an onychophoran, a mollusc and an annelid, and compared them with published gene orders of other species. The result strongly supports the monophyly of Arthropoda and of Mandibulata (atelocerates plus crustaceans) and refutes the Uniramia (atelocerates plus onychophorans). Gene arrangement comparisons are emerging as a powerful new tool for resolving ancient phylogenetic relationships.

Three groups of living arthropods can be recognized with confidence: Atelocerata (insects, myriapods), Crustacea (shrimp, lobsters, barnacles, crabs) and Chelicerata (horseshoe crabs, arachnids). Debates rage, sometimes acrimoniously, on whether these taxa constitute a monophyletic group and on how they are related. Cladistic analyses of morphological characters have generally supported arthropod monophyly¹, but some studies of functional morphology conclude that 'arthropodization' occurred independently in various lineages as each evolved a protective chitinous exoskeleton². Some functional morphologists place onychophorans with atelocerates to form the Uniramia², although evidence from fossil insects suggests that this is not a monophyletic group³. The most closely related arthropod subgroups may be the atelocerates and crustaceans (the Mandibulata); both have mandibles on the fourth head segment and share features of the eye, brain and appendages^{1,4}. However, studies of other aspects of morphology^{5,6} or of fossil evidence and functional considerations^{7,8} unite crustaceans and chelicerates. Comparisons of arthropod ribosomal RNA sequences have yielded conflicting results: some claim arthropod polyphyly⁹; some recognize a monophyletic Arthropoda and Mandibulata, but exclude myriapods from a clade containing the other arthropods^{10,11}; and one unites chelicerates with atelocerates¹². Reasons for conflict include alignment ambiguities, artefactual associations of rapidly evolving taxa, confounding influences of base compositional or substitutional biases, and multiple substitutions at many nucleotide positions¹³⁻¹⁸.

We are examining a set of molecular characters, the arrangement of genes in mitochondrial DNA, that promises to be especially useful for resolving ancient relationships. Metazoan mtDNAs typically encode 36 or 37 genes: 2 for rRNAs, 22 for transfer RNAs, and 12 or 13 for electron transport proteins¹⁹. In some mtDNAs, all genes are transcribed from the same strand; in others, both strands encode genes. The large number of possible gene arrangements makes it improbable that the same order would arise independently. Thus shared-derived arrangements are likely to indicate common ancestry^{13,20-22}.

Complete mitochondrial gene arrangements have been published for 14 invertebrates (3 echinoderms, 4 insects, 1 crustacean, 3 nematodes, 2 molluscs and 1 cnidarian)^{13,19,20,23-25}. The

multiplicity of arrangements found among these groups suggests that particular arrangements are probably not selectively constrained, but rather that rearrangements that do not lethally disrupt mitochondrial genome functioning are rare.

We have determined the relative arrangements of numerous mitochondrial genes for a chelicerate (*Limulus*), two crustaceans (*Homarus* and *Daphnia*), a myriapod (*Thyrophygus*), an onychophoran (*Euperipatoides*), a mollusc (*Plicopurpura*) and an annelid (*Lumbricus*). We compared these and other published mitochondrial gene arrangements cladistically, joining taxa only on the basis of shared-derived characters. As can be seen in Fig. 1, the mtDNAs of *Lumbricus*, the mollusc *Katharina* and the nematodes *Ascaris* and *Caenorhabditis* have in common the direct abutment of two gene pairs; $tRNA^C/tRNA^M$ and $tRNA^{S(AGN)}/ND2$ (ND genes code for NADH dehydrogenase subunits). *Katharina* mtDNA also shares the arrangement $tRNA^W/tRNA^Y$ with another nematode, *Meloidogyne*. In contrast, several arthropods, representing Chelicerata, Crustacea and Atelocerata, share alternative arrangements of these genes: $tRNA^C$ is between $tRNA^W$ and $tRNA^Y$, $tRNA^Y$ is inverted relative to $tRNA^W$, $tRNA^M$ is between $tRNA^Q$ and $ND2$, and $tRNA^{S(AGN)}$ is between $tRNA^N$ and $tRNA^E$ (Fig. 1).

If the arthropod clade shown in Fig. 1 included *Katharina* and/or *Lumbricus*, either the *Katharina* and *Lumbricus* arrangements reverted to those of nematodes, or identical rearrangements occurred independently in two or more arthropod lineages. Neither is likely. The most parsimonious explanation is that the arrangement shared among *Katharina*, *Lumbricus* and nematodes is unchanged from an ancestral state, and the alternative arrangement was derived from it early in the common arthropod lineage. These shared-derived gene arrangements support arthropod monophyly. *Artemia* has apparently acquired the translocation of $tRNA^I/tRNA^Q$ independently; this arrangement is not shared by any organism yet studied.

Similarly, a tRNA gene rearrangement reveals arthropod subgroup relationships. Metazoan mtDNAs contain two tRNA genes specifying leucine, designated $tRNA^{L(CUN)}$ and $tRNA^{L(UUR)}$, according to the codons these tRNAs recognize. In the mtDNAs of *Limulus* and *Euperipatoides* these genes abut directly in the arrangement $l-rRNA/tRNA^{L(CUN)}/tRNA^{L(UUR)}/ND1$. This is identical to their arrangement in *Katharina* and *Plicopurpura*, and similar to that in *Mytilus* (in which $l-rRNA$ is at a different position) and *Lumbricus* (in which there have been independent translocations of $tRNA^A$ and $tRNA^{S(UCN)}$). In contrast, the mtDNAs of the crustaceans *Artemia*, *Daphnia* and *Homarus* share with *Drosophila* the arrangements $COI/tRNA^{L(UUR)}/COII$ (COI and $COII$ code for cytochrome oxidase subunits I and II) and $l-rRNA/tRNA^{L(CUN)}/ND1$, and also share the latter arrangement with *Thyrophygus* (Fig. 1).

It is unlikely that $tRNA^{L(UUR)}$ has translocated to a position between COI and $COII$ independently in the lineages leading to Atelocerata and Crustacea. The more parsimonious explanation is that *Euperipatoides* and *Limulus* retain the ancestral arthropod arrangement, and that the translocation of $tRNA^{L(UUR)}$ to the position between COI and $COII$ occurred in the lineage common to Atelocerata and Crustacea after its separation from that leading to chelicerates and/or onychophorans. If *Thyrophygus* is correctly inferred as an atelocerate, $tRNA^{L(UUR)}$ must have translocated from the $COI/COII$ junction to another position in the mtDNA after myriapods diverged from insects. However, our data do not eliminate the possibility that Atelocerata is not monophyletic, with myriapods diverging earlier than shown in Fig. 1, and $tRNA^{L(UUR)}$ having translocated from the $l-rRNA/ND1$ region convergently with the other mandibulates.

Available data do not allow us to place *Euperipatoides* unambiguously on this phylogenetic tree, but are clearly inconsistent with the Uniramia hypothesis². To place *Euperipatoides* in a clade with the atelocerates would require either independent, identical rearrangements in atelocerates and crustaceans or a

* Present addresses: Department of Cell Biology and Neuroanatomy, University of Minnesota, 4-135 Jackson Hall, 321 Church St. SE, Minneapolis, Minnesota 55455, USA (J.L.B.); and Department of Biology, Eastern Michigan University, Ypsilanti, Michigan 48197, USA (D.S.).

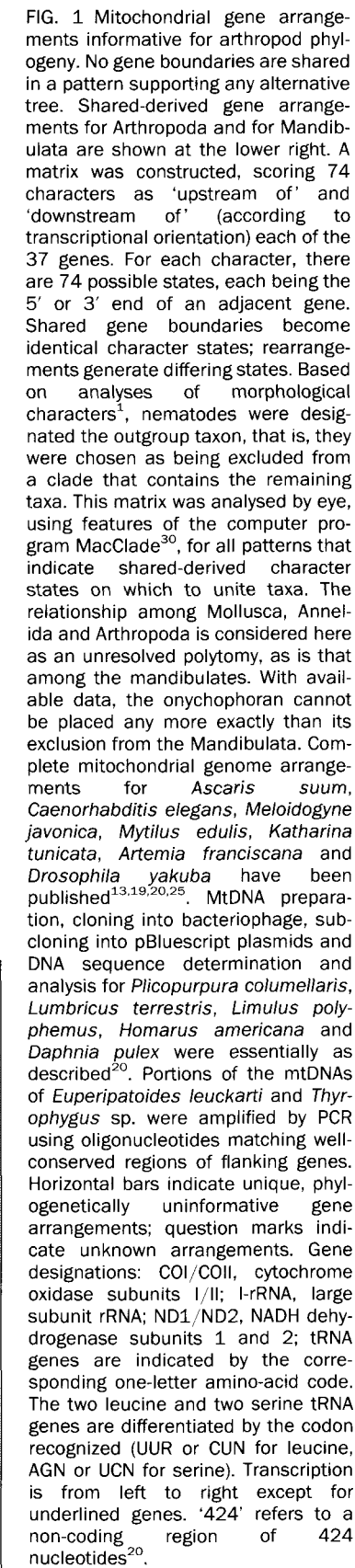


FIG. 1 Mitochondrial gene arrangements informative for arthropod phylogeny. No gene boundaries are shared in a pattern supporting any alternative tree. Shared-derived gene arrangements for Arthropoda and for Mandibulata are shown at the lower right. A matrix was constructed, scoring 74 characters as 'upstream of' and 'downstream of' (according to transcriptional orientation) each of the 37 genes. For each character, there are 74 possible states, each being the 5' or 3' end of an adjacent gene. Shared gene boundaries become identical character states; rearrangements generate differing states. Based on analyses of morphological characters¹, nematodes were designated the outgroup taxon, that is, they were chosen as being excluded from a clade that contains the remaining taxa. This matrix was analysed by eye, using features of the computer program MacClade³⁰, for all patterns that indicate shared-derived character states on which to unite taxa. The relationship among Mollusca, Annelida and Arthropoda is considered here as an unresolved polytomy, as is that among the mandibulates. With available data, the onychophoran cannot be placed any more exactly than its exclusion from the Mandibulata. Complete mitochondrial genome arrangements for *Ascaris suum*, *Caenorhabditis elegans*, *Meloidogyne javonica*, *Mytilus edulis*, *Katharina tunicata*, *Artemia franciscana* and *Drosophila yakuba* have been published^{13,19,20,25}. MtDNA preparation, cloning into bacteriophage, subcloning into pBluescript plasmids and DNA sequence determination and analysis for *Pilcopurpura columellaris*, *Lumbricus terrestris*, *Limulus polyphemus*, *Homarus americanus* and *Daphnia pulex* were essentially as described²⁰. Portions of the mtDNAs of *Euperipatoides leuckarti* and *Thyrophysus* sp. were amplified by PCR using oligonucleotides matching well-conserved regions of flanking genes. Horizontal bars indicate unique, phylogenetically uninformative gene arrangements; question marks indicate unknown arrangements. Gene designations: COI/COII, cytochrome oxidase subunits 1/II; I-rRNA, large subunit rRNA; ND1/ND2, NADH dehydrogenase subunits 1 and 2; tRNA genes are indicated by the corresponding one-letter amino-acid code. The two leucine and two serine tRNA genes are differentiated by the codon recognized (UUR or CUN for leucine, AGN or UCN for serine). Transcription is from left to right except for underlined genes. '424' refers to a non-coding region of 424 nucleotides²⁰.

reversion of the onychophoran mtDNA to the more primitive arrangement of *l-rRNA/tRNA^{L(CUN)}/tRNA^{L(UUR)}/ND1*. Our inference of Onychophora as either the sister group to Chelicerata or to Arthropoda is consistent with the results of comparing mitochondrial rRNA sequences¹¹ and morphological features¹.

The positions of the leucine tRNA genes in other insect mtDNAs are similar or identical to those in *Drosophila*. They are identical in another dipteran, the mosquito *Anopheles*²⁴, and in a hymenopteran, the honeybee *Apis*²³. In an orthopteran, the grasshopper *Locusta*, *tRNA^{L(UUR)}* is upstream of *COII* and *tRNA^{L(CUN)}* is upstream of *ND1*²⁶, although the other flanking genes are currently unknown. A lepidopteran, the moth *Spodoptera*²⁷, and another dipteran, the fly *Simulium*²⁸, have the arrangement *l-rRNA/tRNA^{L(CUN)}/ND1*, but the position of *tRNA^{L(UUR)}* is unknown. Finally, polymerase chain reaction amplification of *COII* from representatives of 10 different insect orders has been achieved using a primer that anneals to *tRNA^{L(UUR)}* (ref. 29). Thus *tRNA^{L(UUR)}* must be immediately upstream of *COII* in each of these insects.

Obviously, not all divergences will coincide with mtDNA arrangements, and independent changes in one or more lineages may subsequently erase traces of relatedness. However, when shared rearrangements are preserved, relationships can be reliably inferred. The great number of arrangements theoretically possible makes it very unlikely that taxa will acquire the same arrangement by chance, so even if some relationships cannot be resolved by such data, those that can are likely to be correct. This is affirmed in this study, where all gene order characters considered are either autapomorphic or support the phylogeny presented here. Although relatively few characters are available for comparison, they are not subject to many of the shortcomings attributed to morphological studies (for example, convergence resulting from common selective pressures, or ambiguity in determining homologous structures), functional arguments (for example, lack of falsifiability), or primary sequence comparisons (for example, problems of alignment, base compositional effects, or excessive homoplasy). The relationships of taxa thought to share a close evolutionary history with arthropods, such as tardigrades and pentastomids, can be addressed with similar data. We anticipate that, as mitochondrial gene arrangements are determined for more taxa, additional evolutionary relationships will be clarified. □

ACKNOWLEDGEMENTS. This work was supported by grants from the NIH, NSF and the Rackham Graduate School of the University of Michigan. We thank E. Weinstein and B. Salisbury for technical assistance; J. M. Turbeville for the gift of *Euperipatoides* mtDNA and PCR primers; and S. I. Fuerstenberg, J. M. Turbeville, F. Kraus and G. Naylor for comments. PCR primer sequences and the matrix used for comparing gene arrangements are available from the authors. All DNA sequences except those for *Homarus* and *Daphnia* have been placed in GenBank under accession numbers U24570 and U29703–U29719; *Homarus* and *Daphnia* sequences are available on request.

Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods

Markus Friedrich & Diethard Tautz

Zoologisches Institut der Universität München, Luisenstrasse 14, 80333 München, Germany

THE evolutionary relationships among arthropods are of particular interest because the best-studied model system for ontogenetic pattern formation, the insect *Drosophila*, is a member of this phylum. Evolutionary inferences about the developmental mechanisms that have led to the various designs of the arthropod body plan depend on a knowledge of the phylogenetic framework of arthropod evolution. Based on morphological evidence^{1–3}, but also on palaeontological considerations⁴, the sister group of the insects is believed to be found among the myriapods. Using nuclear ribosomal gene sequences for constructing a molecular phylogeny, we provide strong evidence that the crustaceans and not the myriapods should be considered to be the sister group of the insects. Moreover, the degree of sequence divergence suggests that the diversification of the myriapods occurred during the Cambrian. Our findings have general implications for the course of land colonization by the different arthropod groups, as well as for the interpretation of primitive and derived features of arthropod morphology.

To study the phylogenetic relationships of arthropods, we have obtained extensive sequence information from the nuclear ribosomal genes of taxa that cover most of the evolutionary divergence of each of the four major extant arthropod subgroups (Fig. 1). The data were analysed with respect to parameters that are known to be crucial for the accuracy of molecular phylogenetic methods such as homogeneity of nucleotide composition and similarity of evolutionary rates (Fig. 1). To estimate phylogenetic trees, we applied neighbour-joining with distances corrected for multiple hits and gamma-distributed rates across sites⁵, weighted maximum parsimony⁶ and maximum likelihood⁷. The bootstrap method was used to quantify the support for the tree topology⁸.

We find that all three procedures agree with very high support on three major nodes (nodes *a*, *b* and *c* in Fig. 1). Node *a* supports the monophyletic origin of the arthropod taxa included, node *b* concerns the monophyly of the Chelicerata and node *c* unites the Crustacea and the Hexapoda into a monophyletic group. These results have also been partly suggested previously on the basis of less extensive comparisons of both nuclear and mitochondrial ribosomal genes of some relevant taxa (reviewed in refs 3, 9).

The major groups reconstructed in our tree correspond in principle to the concept of the four extant arthropod classes: Myriapoda, Chelicerata, Crustacea and Hexapoda. The support for the monophyly of the Myriapoda is also notable, given the numerous suggestions for their being paraphyletic with respect to the Hexapoda^{3,10}. But the support for the monophyletic status of the Crustacea is weak. Our data suggest that these might also be paraphyletic with respect to the Insecta. Whereas the shortest

Received 11 October 1994; accepted 18 May 1995.

1. Brusca, R. C. & Brusca, G. L. *Invertebrates* (Sinauer, Sunderland, Massachusetts, 1990).
2. Manton, S. M. & Anderson, D. T. in *The Origin of Major Invertebrate Groups* (ed. House, M. R.) 269–322 (Systematics Association, London, 1979).
3. Kukalová-Peck, J. *Can. J. Zool.* **70**, 236–255 (1992).
4. Boudreaux, H. B. *Arthropod Phylogeny with Special Reference to Insects* (Wiley, New York, 1979).
5. Tiegs, O. W. Q. *J. microsc. Sci.* **82**, 165–268 (1947).
6. Cisse, J. L. *Science* **186**, 13–18 (1974).
7. Hessler, R. R. & Newman, W. A. *Fossils Strata* **4**, 437–459 (1975).
8. Valentine, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2272–2275 (1989).
9. Lake, J. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 763–766 (1990).
10. Field, K. G. *et al. Science* **239**, 748–753 (1988).
11. Ballard, J. W. O. *et al. Science* **258**, 1345–1348 (1992).
12. Hendriks, L., van Broeckhoven, C., Vandenbergh, A., van de Peer, Y. & DeWachter, R. *Eur. J. Biochem.* **177**, 15–20 (1988).
13. Boore, J. L. & Brown, W. M. *Nautilus* **108** (Suppl. 2), 61–78 (1994).
14. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. *Cladistics* **9**, 1–39 (1993).
15. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
16. Vawter, L. & Brown, W. M. *Genetics* **134**(2), 597–608 (1993).
17. Patterson, C., Williams, D. M. & Humphries, C. J. *A. Rev. Ecol. Syst.* **24**, 153–188 (1993).
18. Adoutte, A. & Philippe, H. *Experientia* (Suppl.) **63**, 1–30 (1993).
19. Wolstenholme, D. R. *Int. Rev. Cytol.* **141**, 173–216 (1992).
20. Boore, J. L. & Brown, W. M. *Genetics* **138**, 423–443 (1994).
21. Sankoff, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 6575–6579 (1992).
22. Smith, M. J., Arndt, A., Gorski, S. & Fajber, E. *J. molec. Evol.* **36**, 545–554 (1993).
23. Crozier, R. H. & Crozier, Y. C. *Genetics* **133**, 97–117 (1993).
24. Mitchell, S. E., Cockburn, A. F. & Seawright, J. A. *Genome* **36**, 1058–1073 (1993).
25. Valverde, J. R., Batuecas, B., Moratilla, C., Marco, R. & Garesse, R. *J. molec. Evol.* **39**, 400–408 (1994).
26. Haucke, H.-R. & Gellissen, G. *Curr. Genet.* **14**, 471–476 (1988).
27. Pashley, D. P. & Ke, L. D. *Molec. Biol. Evol.* **9**(6), 1061–1075 (1992).
28. Pruess, K., Zhu, X. & Powers, T. J. *med. Ent.* **29**(4), 644–651 (1992).
29. Liu, H. & Beckenbach, A. T. *Molec. Phylog. Evol.* **1**, 41–52 (1992).
30. Maddison, W. P. & Maddison, D. R. *Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, Massachusetts, 1992).

maximum-parsimony tree and the best maximum-likelihood tree favour monophyly of the Crustacea, in the bootstrap analyses with these methods, *Artemia salina* is positioned closer to the Insecta (not shown). Neighbour-joining does not sufficiently differentiate this part of the tree. Finally, all trees favour a sister-group relationship between Myriapoda and Chelicerata hitherto not considered, although the support is not at a significant level.

The Chelicerata and Myriapoda species show similar molecular substitution rates and their terminal branch lengths can thus be compared directly (Fig. 1). This comparison suggests that the two major diplopod taxa split at about the same time as the Chelicerata, that is, probably in the Late Cambrian¹¹. Given the poor fossil documentation of Myriapoda, which reaches back to the Late Silurian^{12,13}, this is a significant finding. It is probable that, like the chilopods, all other extant major myriapod lineages already existed when diplopods diversified¹⁰. It can therefore be concluded that the major subgroups of the Myriapoda split off during the Cambrian. As animal life on land was not possible before the Ordovician¹³, land colonization must have been achieved independently by the major myriapod lineages. This may also explain the differences of tracheal organization among major myriapodal subgroups¹⁰.

In view of the possibility of multiple independent land colonizations, one has to re-evaluate all character states for potential convergences that could have been caused by the adaptive constraints posed by the transition from water to land. This concerns in particular those character states that have been used to unite the land-living classes Insecta and Myriapoda into a higher taxon called Tracheata, Antennata or Atelocerata^{1,2}, that is, the possession of Malpighian tubules, a tracheal system and one pair of antennae only.

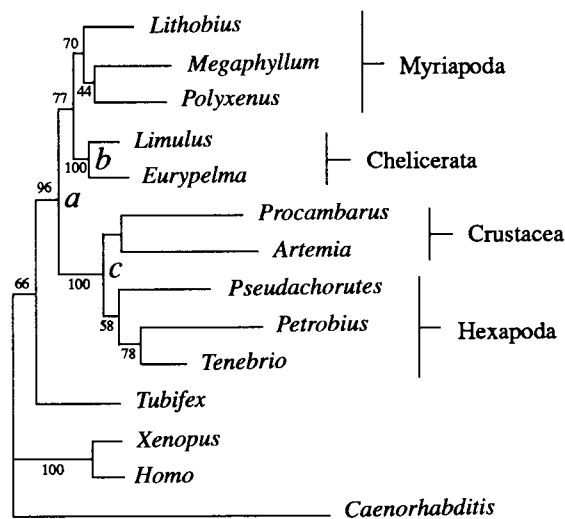


FIG. 1 Phylogenetic relationships of the arthropod taxa as estimated from 18S and 28S rDNA sequences. Insecta are represented by an endognath spring-tail (*Pseudachorutes* sp., Collembola), and two ectognath species, an apterous bristle-tail (*Petrobius brevistylis*, Archaeognatha) and a beetle (*Tenebrio molitor*, Holometabola, Coleoptera). Myriapoda are represented by a centipede (*Lithobius forficatus*, Chilopoda) and two most distantly related millipede, i.e. diplopod, species (*Megaphyllum* sp., Chilognatha; *Polyxenus lagurus*, Pselaphognatha). The Crustacea are represented by a brine shrimp (*Artemia salina*, Branchiopoda) and a crayfish (*Procamburus clarkii*, Eumalacostraca), and the Chelicerata by a spider (*Eurypelma californica*, Arachnida) and a horseshoe crab (*Limulus polyphemus*, Xiphosura). The outgroup taxa included are *Tubifex* sp. (Annelida), *Xenopus laevis* (Amphibia), *Homo sapiens* (Mammalia) and *Caenorhabditis elegans* (Nematoda). The tree topology and branch lengths were estimated with the maximum-likelihood algorithm²² assuming a transition transversion ratio of 2. The bar below the tree corresponds to estimated 0.05

Malpighian tubules are the major excretory organs in Myriapoda and Insecta. Functionally they replace excretion through gills in water-living Arthropods. As with tracheal respiration, such organs are also found in many land-living arachnid taxa¹⁴. Thus, despite the similarities in their ontogenetic origins¹⁰, the commonly assumed homology of the Malpighian tubules in Insecta and Myriapoda is not firmly established. Another character state that is considered to be synapomorphic for the Tracheata is the lack of appendages on the third head segment^{2,10}. The ancestral function of the second antenna in the arthropod ground plan is that of stridulatory locomotion under water and of filter feeding^{4,14}. On land, this function is dispensable; consequently, the second antenna is unnecessary and may thus have been lost independently in Myriapoda and Insecta.

Interestingly, apart from the molecular evidence presented here, there is also morphological evidence in support of a sister-group relationship between insects and crustaceans. One concerns the structure of the ommatidia, which is highly conserved in these groups¹⁵, whereas the ommatidia of extant Chelicerata and Myriapoda look more primitive. In the Myriapoda, this has been interpreted to be the result of a secondary modification¹⁵, but could also be considered as an ancestral state. Further, some aspects of the development of the central nervous system appear strikingly conserved between Insecta and Crustacea. In representatives of these arthropod classes, neurogenic stem cells produce ganglion mother cells by unequal cell divisions in dorso-ventral orientation^{16,17}. In Myriapoda, the neuroectoderm produces clusters of ganglion mother cells direct by randomly oriented cell division, which invaginate in a dorsal direction¹⁶⁻¹⁸. As the latter situation is also found in Chelicerata and Onychophora^{16,17}, one could consider the characteristics of

substitutions per site. Numbers at the branches correspond to the percentage bootstrap support from 100 replications with the maximum-likelihood algorithm. No number is given for the branch leading to the crustacean taxa. The reason for this is that, in contrast to the situation shown, the bootstrap analysis tentatively (65%) favours crustacean paraphyly with *Artemia salina* being sister group to the Insecta (see text). The topology shown was also obtained by using maximum-parsimony⁶ weighting transversions as twice transitions. Neighbour-joining with distances corrected for multiple hits, taking a gamma distribution of site-to-site substitution probability with $\alpha=1$ into account⁵, also reconstructed the tree shown, except that the relationship between Insecta and Crustacea was not resolved. Bootstrap analysis with 1000 replications of the neighbour-joining or the maximum-parsimony tree yielded comparable support for all other branches. It is clear that under extreme conditions of rate heterogeneity, all molecular tree estimation methods tend to unite fast taxa artificially^{24,25}. According to the relative rate test²⁶ corrected for multiple testing of 43 pairwise in-group taxon combinations ($z=3.35$ for $P=0.05$) and using *Tubifex* sp. as outgroup, myriapod and chelicerate lineages evolved at homogeneous rates, but most insect and crustacean lineages significantly faster (2–3-fold). The beetle (*Tenebrio molitor*), however, represents a slow-clock species among the insects. For similar reasons we omitted *Drosophila melanogaster* from the tree presented. The ribosomal genes of this species show extremely accelerated evolutionary rates. When included in our analysis, only maximum likelihood places *Drosophila* next to the beetle, whereas all other methods fail to place it within the Hexapoda (not shown). Thus the relative robustness of the maximum-likelihood method against rate heterogeneity among taxa²⁷ can be confirmed by our data. Compositional bias was also considered as a possible source of error. According to χ^2 statistics the 814 variable sites of the whole data set show no significant nucleotide heterogeneity among the in-group taxa ($\chi^2=26$, d.f.=30) (ref. 28).

METHODS. Five different DNA fragments corresponding to nucleotides 20–1241, 1389–1830, 3319–3677, 4046–4746 and 4995–5482 of the *Drosophila melanogaster* rRNA sequence²⁹ were amplified by polymerase chain reaction and directly sequenced in both directions. The sequences were aligned crudely using ClustalV³⁰ and optimized, taking secondary structure into account. EMBL accession numbers of sequences, alignment and primer sequences are available on request (e-mail: tautz@zi.biologie.uni-muenchen.de). Ambiguous regions and all sites containing gaps in one of the in-group taxa were removed, yielding 1067 sites from the 18S and 786 sites from the 28S rDNA genes for subsequent analysis.

central nervous system development in Insecta and Crustacea as a derived condition. Similarly, the axonal precursor cells show a greater evolutionary conservation between Insecta and Crustacea than between Insecta and Myriapoda¹⁹.

Another often-disputed character state is the first pair of gnathal appendages, the mandibles. Their functional specification is considered to be of common origin in Myriapoda, Insecta and Crustacea, which may therefore be united as Mandibulata¹. Whereas in the latter two classes the mandibles are short, solid and gnathobasic, there are examples of jointed mandibles in every major subgroup of the Myriapoda²⁰. This situation has been interpreted to be either ancestral²⁰ or the consequence of secondary articulation of a gnathobasic myriapod mandible²¹. The molecular phylogenetic evidence suggests that the jointed mandibles of the Myriapoda may represent an ancestral state, that is, a less modified arthropod limb. Independent support comes from a Cambrian myriapod-like fossil, which is described to have possessed long jointed appendages instead of solid mandibles²².

Given the conflict between the classical Tracheata concept and the results of the molecular phylogenetic studies, future investigations should also include a search for phylogenetic informative character states at the molecular level. In *Drosophila*, many genes have been characterized, which are involved in the development of the structures discussed above. They may be used as molecular probes to test homology in other arthropod classes. The homeobox gene *distalless*, for example, specifies the distal portion of appendages, but is not expressed in the mandibular segment of insects confirming the gnathobasic character of the insect mandible²³. On the basis of the data presented above we would suggest that *distalless* should still be expressed in the mandible primordia of those myriapodan taxa that bear jointed mandibles. □

Embryonic lethality and liver degeneration in mice lacking the RelA component of NF- κ B

Amer A. Beg*, William C. Sha*,
Roderick T. Bronson†, Sankar Ghosh‡
& David Baltimore*§

* Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Department of Pathology, Tufts University School of Medicine and Veterinary Medicine, Boston, Massachusetts 02111, USA

‡ Section of Immunobiology and the Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06520-8011, USA

NF- κ B, which consists of two polypeptides, p50 (*M_r* 50K) and p65/RelA (*M_r* 65K), is thought to be a key regulator of genes involved in responses to infection, inflammation and stress¹. Indeed, although developmentally normal, mice deficient in p50 display functional defects in immune responses². Here we describe the generation of mice deficient in the RelA subunit of NF- κ B. Disruption of the *relA* locus leads to embryonic lethality at 15–16 days of gestation, concomitant with a massive degeneration of the liver by programmed cell death or apoptosis. Embryonic fibroblasts from RelA-deficient mice are defective in the tumour necrosis factor (TNF)-mediated induction of messenger RNAs for I κ B α and granulocyte/macrophage colony stimulating factor (GM-CSF), although basal levels of these transcripts are unaltered. These results indicate that RelA controls inducible, but not basal, transcription in NF- κ B-regulated pathways.

The gene for the RelA subunit of NF- κ B was disrupted in embryonic stem cells by homologous recombination (Fig. 1a). Chimaeric mice generated heterozygous animals, which were interbred. No phenotypic abnormalities were observed in the heterozygous mice but, out of a total of 50 pups genotyped from these matings, no homozygotes were obtained (12.5 expected), suggesting that RelA is essential for survival. Timed matings were subsequently used to obtain embryos at different stages of development. Up to 14 days of gestation, wild-type, heterozygous and homozygous animals were present in the expected mendelian ratios (Fig. 1b). Embryonic fibroblasts were obtained from 13-day embryos and analysed by western blotting using antibodies against RelA (Fig. 1c). RelA was absent in homozygous fibroblasts, and was present in smaller amounts in heterozygous fibroblasts.

Among 16-day embryos, typically one-quarter were dead. Genotyping showed that the dead embryos were RelA knockouts. Embryos that had not been extensively resorbed showed a massive haemorrhage in the abdominal cavity. Viable 15–16-day RelA knockouts as well as wild-type embryos were analysed histologically (Fig. 2). Sections of the RelA knockouts showed that a massive degeneration of the liver had taken place (Fig. 2b). Areas of tissue loss were apparent in many parts of the liver. In addition, it seemed that hepatocytes were being specifically targeted, whereas haematopoietic precursors and red blood cells remained intact. In fact, certain areas of the liver from knockouts showed potentially enhanced erythropoiesis (not shown). The abdominal haemorrhage seen in moribund embryos was thus probably due to the breakdown of liver architecture and may have closely preceded death of the embryos. No other developmental abnormalities were noticed in the 15-day knockout embryos. Moreover, liver sections of 14-day-old RelA knockouts were indistinguishable from those of wild-type or heterozygous

Received 3 February; accepted 2 June 1995.

- Weygoldt, P. Zool. Syst. Evol.-forsch. **24**, 19–35 (1986).
- Kristensen, N. P. In *The Insects of Australia* 2nd edn (eds Naumann, I. D. et al.) 125–140 (Melbourne Univ. Press, 1991).
- Štys, P. & Zrzavý, J. Eur. J. Ent. **91**, 257–275 (1994).
- Lauterbach, K.-E. Abh. naturw. Ver. Hamburg (NF) **23**, 105–161 (1980).
- Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01 (Pennsylvania State Univ., 1993).
- Swofford, D. L. *Phylogenetic Analysis Using Parsimony*, 3.4.1 (Illinois Nat. Hist. Surv., Champaign, 1993).
- Felsenstein, J. *Phylogenetic Inference Programs (PHYLIP)* (Univ. of Washington, 1989).
- Felsenstein, J. *Evolution* **39**, 791–793 (1985).
- Averoff, M., Akam, M. Phil. Trans. R. Soc. **B347**, 293–303 (1995).
- Dohle, W. Abh. naturw. Ver. Hamburg (NF) **23**, 45–104 (1980).
- Robison, R. & Kaesler, R. L. in *Fossil Invertebrates* (eds Boardman, R. S., Cheetham, A. H. & Rowell, A. J.) 206–220 (Blackwell, Oxford, 1987).
- Jeram, A. J., Selden, P. A. & Edwards, D. Science **250**, 685–661 (1990).
- Kukalová-Peck, J. in *The Insects of Australia* 2nd edn (eds Naumann, I. D. et al.) 141–179 (Melbourne Univ. Press, 1991).
- Brusca, C. & Brusca, G. J. *Invertebrates* (ed. Sinauer, A. D.) (Sinauer Associates, Sunderland, MA, 1990).
- Paulus, H. F. in *Arthropod Phylogeny* (ed. Gupta, A. P.) 299–371 (Van Nostrand, New York, 1979).
- Weygoldt, P. in *Neurobiology of Arachnids* (ed. Barth, F. G.) 20–37 (Springer, New York, 1985).
- Anderson, D. T. *Embryology and Phylogeny in Annelids and Arthropods* (Pergamon, Oxford, 1973).
- Whittington, P. M., Meier, T. & King, P. Wilhelm Roux Arch. dev. Biol. **199**, 349–363 (1991).
- Whittington, P. M., Leach, D. & Sandeman, R. Development **118**, 449–461 (1993).
- Manton, S. M. Phil. Trans. R. Soc. **B247**, 1–183 (1964).
- Boudreaux, H. B. *Arthropod Phylogeny* (ed. Gupta, A. P.) 551–584 (Van Nostrand, New York, 1979).
- Robison, R. A. Nature **343**, 163–164 (1990).
- Panganiban, G., Nagy, L. & Carroll, S. B. Curr. Biol. **4**, 671–675 (1994).
- Huelsenbeck, J. P. & Hillis, D. M. Syst. Zool. **42**, 247–264 (1993).
- Gaut, B. S. & Lewis, P. O. Molec. Biol. Evol. **12**, 152–162 (1995).
- Wu, C.-I. & Li, W.-H. Proc. natn. Acad. Sci. U.S.A. **82**, 1741–1745 (1985).
- Hillis, D. M., Huelsenbeck, J. P. & Swofford, D. L. Nature **369**, 363–364 (1994).
- Weir, B. S. Genetic Data Analysis (Sinauer Associates, Sunderland, MA, 1990).
- Tautz, D., Hancock, J. M., Webb, D. A., Tautz, C. & Dover, G. Molec. Biol. Evol. **5**, 366–376 (1988).
- Higgins, D. G. & Sharp, P. M. Gene **73**, 237–244 (1988).

ACKNOWLEDGEMENTS. We thank S. Benzer, G. Boyan, B. Hausdorf, R. Melzer, A. Minelli and A. von Haeseler for suggestions on the manuscript; P. Makings for providing samples for the study; A. Minelli for help with the classification of the myriapod and the collembolan specimens; and B. Hausdorf for making part of the sequences of *Eurypelma californica* accessible before publication. The work was supported by a Ph.D. studentship of the University of Munich to M.F. and by a Gerhard Hess prize of the DFG to D.T.

§ To whom correspondence should be addressed.

embryos (not shown). These results indicate that RelA is not required for earlier steps in embryogenesis.

Liver sections of RelA knockouts contained areas where nuclear fragmentation had taken place (visible as pyknotic nuclei). Nuclear fragmentation is typically associated with programmed cell death or apoptosis³. To determine whether cell death was due to apoptosis, we analysed embryo sections by the TUNEL assay⁴ (Fig. 2c-f). Liver sections from 15–16-day-old wild-type embryos showed a few isolated cells that were TUNEL-positive. However, the RelA-deficient embryos showed large areas containing numerous TUNEL-positive cells. Thus, cell death in livers of RelA deficient embryos was taking place largely, if not exclusively, by apoptosis. Other regions of the knockouts did not show enhanced apoptosis. In fact, apoptosis in the central nervous system was observed to the same extent in both wild-type and RelA knockout embryos (not shown).

RelA-containing complexes are inducible by a variety of different agents⁵. Among these is TNF α , which can activate NF- κ B in a variety of cell types. To determine whether TNF α could activate κ B-binding activity in the RelA-deficient animals we treated embryonic fibroblasts with TNF α for different lengths of time, after which nuclear extracts were obtained (Fig. 3).

Wild-type fibroblasts showed a dramatic increase in κ B-binding activity after treatment with TNF α (Fig. 3a, lanes 1–3). In contrast, RelA-deficient fibroblasts showed very little inducible binding activity (Fig. 3a, lanes 4–6). The TNF α -inducible binding activity in wild-type fibroblasts was reactive to both p50 and RelA antibodies, whereas the activity in the RelA-deficient fibroblasts was reactive only to p50 antibodies (Fig. 3b, lanes 1–6).

The physiological consequence of the deficiency in TNF α responsiveness in the knockout fibroblasts was determined by the examination of mRNA levels of I κ B α and GM-CSF, two genes that are TNF α responsive and contain NF- κ B sites¹. Levels of mRNAs of both genes were increased after TNF α treatment of wild-type fibroblasts (Fig. 3c). However, no appreciable increase was noticed after TNF α treatment of RelA knockout fibroblasts. Interestingly, no difference was observed in the basal levels of expression of these two genes, suggesting that RelA is involved in the regulation of inducible but not the basal transcription of these genes. These results suggest that RelA is a key regulator of genes inducible by TNF α .

Members of the NF- κ B family are typically present in the cytoplasm bound to the inhibitory I κ B proteins, which include

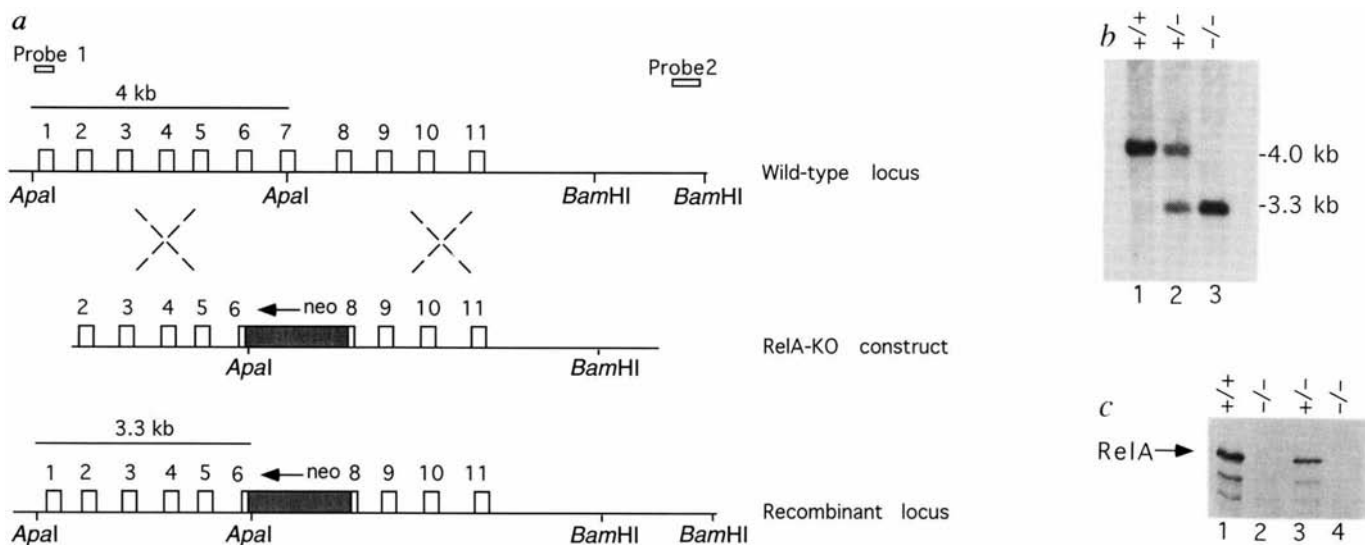


FIG. 1 Generation of mice deficient in the RelA subunit of NF- κ B. **a**, The intron-exon structure of the murine RelA gene (top). Exons are indicated by open boxes. The probes and restriction enzyme sites used for diagnostic Southern blotting are also indicated. The targeting construct (middle), the areas predicted to undergo homologous recombination (crosses), and the recombinant locus (bottom) are also shown. As predicted from the restriction map of the wild-type locus, digestion of genomic DNA with *Apal* generates a 4-kb fragment, which is replaced by a 3.3-kb fragment in the mutant as indicated. **b**, Analysis of DNA preparations from heads of 14-day embryos from heterozygote matings. Southern blot analysis of the 3 genotype categories with probe 1 is shown with wild-type, heterozygote and mutant (knockout) mice, indicated as (+/+), (+/-) and (-/-), respectively. As expected, a *Bam*HI digest followed by hybridization to probe 2 generated the same-size fragment in all 3 genotypes (not shown). **c**, Western blot analysis of embryonic fibroblast extracts with RelA-specific antibodies¹⁸. Extracts from the 3 genotype categories were used as indicated above the lanes. The mobility of the RelA polypeptide is indicated by an arrow.

METHODS. The RelA targeting vector was generated with the pPNT plasmid¹⁹. The complementary DNA sequence of RelA (ref. 20) was used to isolate genomic clones from a 129/Sv mouse liver genomic library (Stratagene). The upstream genomic fragment was excised as a 2.1-kb *Bsm*I fragment and inserted into the *Eco*RI site of pPNT in the

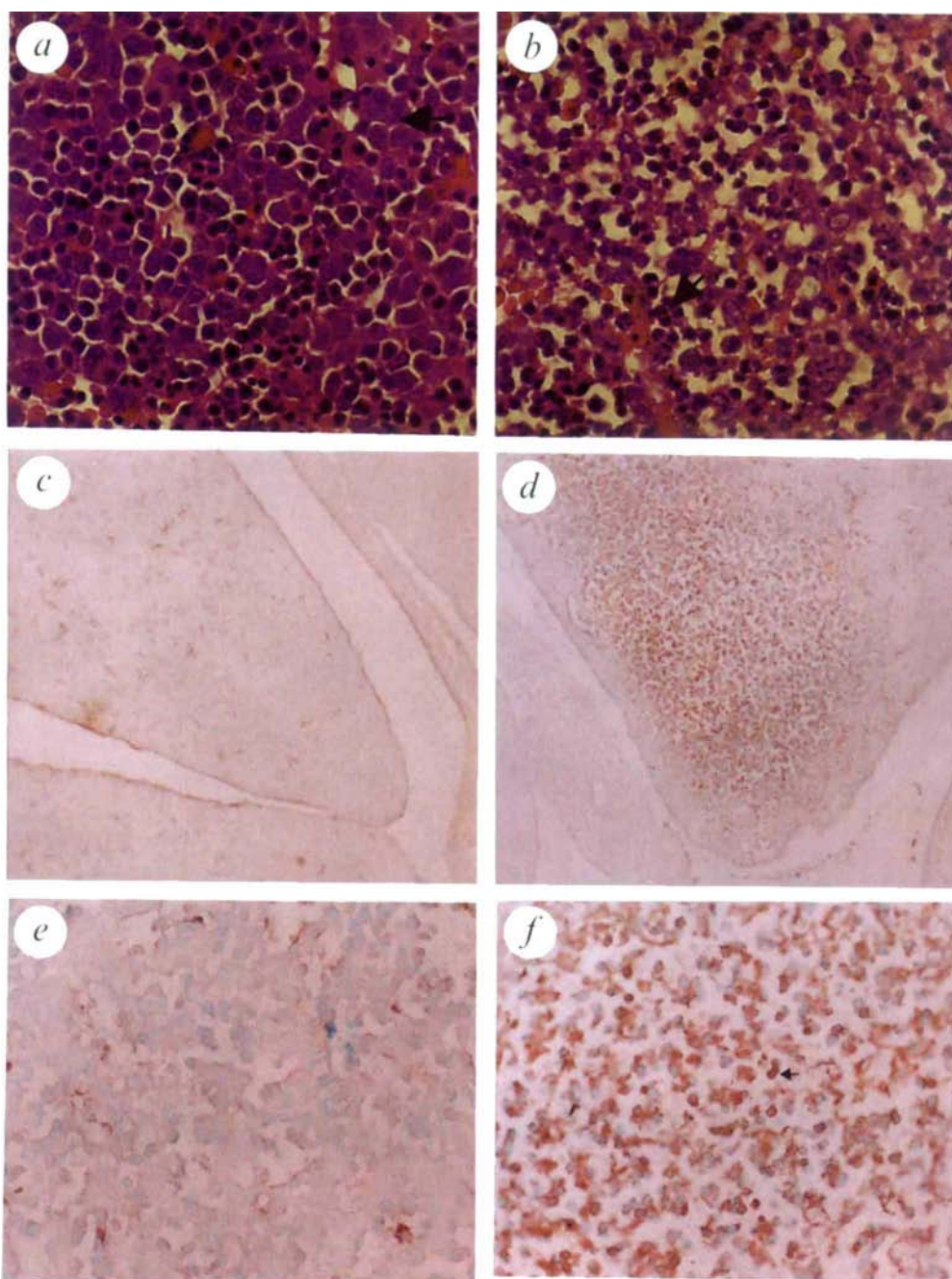
opposite transcriptional orientation from the *pgk-neo* gene. This fragment contains sequences from codons 2 to 155 of RelA (present in exons 2 to 6). The downstream fragment was excised as a *Sac*I fragment from the DNA library and cloned into the *Not*I site of pPNT. This fragment starts at codon 287 in exon 8 and contains 12 kb of downstream sequence. Homologous recombination with this construct (pPNT-RelAKO) would lead to truncation of the RelA polypeptide at residue 155, thereby deleting a significant portion of the Rel-homology region (300 residues) required for DNA binding, dimerization and interaction with I κ B proteins¹. pPNT-RelAKO was made linear at a unique *Sall* site and electroporated into J1 embryonic stem cells (provided by R. Jaenisch), after which G418 and GANC-resistant colonies were selected as previously described²¹. Colonies that had undergone homologous recombination were identified by Southern blotting. Positive clones were injected into C57BL6/J blastocysts. Male chimaeric mice with >70% coat colour chimaerism gave germline transmission when crossed with C57BL6/J females. DNA preparations from the tails of the offspring of these matings were analysed by Southern blots to determine their genotype. Heterozygotes from these matings were interbred to obtain homozygotes. Embryonic fibroblasts were obtained from 13-day embryos. Cell extracts were obtained by lysis of fibroblasts in a buffer containing 0.3% NP-40 (ref. 18). Western blot analyses of these extracts were done with a C-terminal RelA antibody (provided by A. Baldwin)¹⁸.

I κ B α (ref. 6) and I κ B β (ref. 7). Activation of NF- κ B involves the signal-induced degradation of these proteins, allowing NF- κ B to translocate to the nucleus⁸. Interestingly, RelA has been shown to regulate the mRNA expression and stability of the I κ B α protein⁹. As shown above, RelA regulates inducible, but not basal, I κ B α expression. To determine the effect of RelA on the stability of I κ B α we did a western blot analysis with I κ B α -specific antibodies¹⁰. A slight reduction in I κ B α levels was noticed in the RelA knockout fibroblasts (Fig. 4a). In contrast, western blot analysis with I κ B β -specific antibodies⁷ showed no detectable protein in the RelA-deficient fibroblasts (Fig. 4a). The absence of I κ B β in the knockout is probably due to decreased stability in the absence of RelA because I κ B β mRNA levels were unchanged (Fig. 4b). Moreover, I κ B β expression was not inducible, as previously documented⁷. These results show that RelA may regulate the I κ B proteins by controlling both their

inducible expression (I κ B α) and their stability (I κ B α and I κ B β).

The p50/RelA heterodimer (NF- κ B) is the most widely distributed κ B-binding activity¹. Our previous studies on p50-deficient mice revealed multiple defects in immune responses². However, unlike the RelA-deficient mice, the p50 knockouts were developmentally normal. In addition, the recently described RelB-deficient mice, although defective in the production of dendritic cells, do not have major developmental abnormalities^{11,12}. Thus, RelA may have unique functions that do not require p50 or RelB. Such functions may be performed through association of RelA with members of the Rel or non-Rel transcription factors. Previous studies have indicated a role for the c-Rel protein in both cell growth and apoptosis¹³. c-Rel, by virtue of its Rel-homology region, associates with other members of the NF- κ B family, including RelA (ref. 1). Such heterodimeric complexes may interact with critical regulatory elements of genes involved

FIG. 2 Histological analysis of 15.5-day wild-type (WT) and RelA knockout (KO) embryos. *a, b*, Haematoxylin–eosin-stained sagittal liver sections from WT and KO embryos. Hepatocytes present in the WT sections (arrow) appear as lightly stained cells, whereas the haematopoietic precursors appear as the more darkly stained cells. Pycnotic nuclei are arrowed in *b*. TUNEL staining of liver sections from WT and KO embryos is shown both at lower magnification (*c, d*) and at higher magnification (*e, f*). TUNEL-positive cells were rare in WT sections but frequent in KO sections (arrowed in *f*). For *d* and *f*, a RelA KO embryo that showed only partial liver degeneration was used because more TUNEL-positive cells are visible at this stage than at latter stages when few intact cells are present. **METHODS.** Whole embryos were fixed in buffered formalin and embedded in paraffin. Sagittal sections were cut and stained with haematoxylin–eosin. Unstained sections were used for the TUNEL assay, which was done as described⁴.



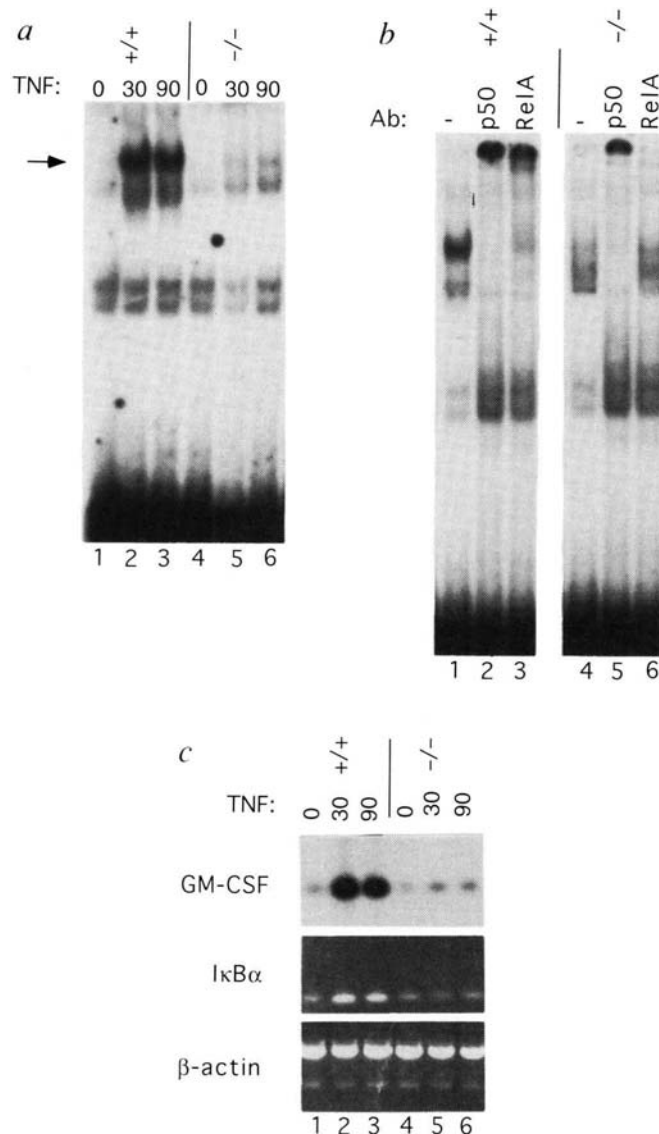


FIG. 3 TNF α responsiveness of WT and RelA KO embryonic fibroblasts. **a**, Electrophoretic mobility shift analysis of nuclear extracts from WT and KO EFs treated with TNF α for the indicated times (in minutes) (lanes 1–6). The major TNF α -inducible κ B-site binding activity is arrowed. **b**, Nuclear extracts obtained from cells treated with TNF α for 90 min submitted to treatment by anti-p50 or anti-RelA antibody. Lanes 1 and 4 are controls. **c**, RT-PCR analysis of mRNAs from WT and RelA KO embryonic fibroblasts. Specific primers were used to amplify regions of cDNA copied from total RNA. The probes were designed from sequences of genes shown at the left. The times (minutes) for which the cells were exposed to TNF α are shown at the top.

METHODS. **a**, **b**, Nuclear extracts were prepared as described¹⁸. Mobility shift assays were done with a hairpin H-2 probe²². The p50 and RelA antibodies have been described^{18,22}. **c**, Total RNA was isolated with the Tri-Reagent (Molecular Research Center). cDNA was made with Stratascript (MMTV-RT) from Stratagene. cDNAs obtained were tested at various concentrations to determine the linear range for the PCR reactions, which were done as follows: (1) 1 cycle at 94 °C for 5 min, 60 °C for 1 min and 72 °C for 1.5 min; (2) 30 cycles at 94 °C for 45 s, 60 °C for 1 min and 72 °C for 1.5 min; (3) 1 cycle at 94 °C for 45 s, 60 °C for 1 min and 72 °C for 10 min. The samples were analysed on 2% agarose gels. Sequences for the GM-CSF and β -actin primers have been described²³. Products synthesized with GM-CSF primers were hybridized to an internal oligonucleotide after Southern transfer²³. The sequences of I κ B α primers used for PCR were from the mouse cDNA sequence (provided by I. Verma). The sequences for these primers were: (1) 5'-CAGCCCCGACAGCCATGTTTCAG, (2) 5'-CATGGAGTCCAGGCCGTGTCGTG. Identical results were obtained when the RT-PCR analysis was done with independently isolated EFs.

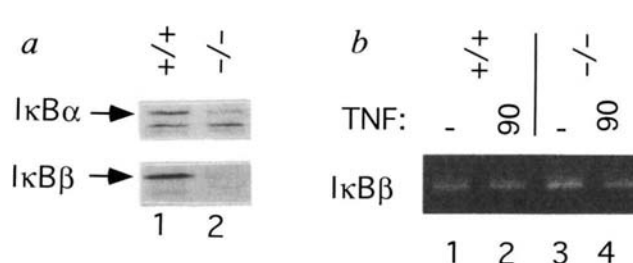


FIG. 4 Post-transcriptional regulation of I κ B proteins by RelA. **a**, Western blot analysis of I κ B α and I κ B β proteins. Cytoplasmic extracts from WT or RelA KO EFs were used as indicated. The mobility of the two proteins is indicated by arrows. **b**, RT-PCR analysis of I κ B β mRNA. RNA obtained from untreated (lanes 1 and 3) or TNF α -treated (90 min; lanes 2 and 4) WT or RelA KO EFs was analysed by RT-PCR using I κ B β -specific primers.

METHODS. Western blot analysis was done as described¹⁸. Both the I κ B α (provided by N. Rice) and I κ B β antibodies have been described^{7,10}. RT-PCR was done as described above. I κ B β -specific primers⁷ used were: (1) 5'-GGGCTCTCTAGTCC, (2) 5'-GGGGCATGGCTGGTG.

in the control of cell proliferation and apoptosis. NF- κ B has also been shown to increase the transcriptional activity of other transcription factors¹⁴, including c-Jun (ref. 15). Interestingly, disruption of c-Jun also leads to embryonic lethality and liver defects¹⁶. Both RelA and c-Jun could be required for the regulation of key genes involved in hepatocyte survival.

It is currently unclear whether the absence of RelA directly causes cell death in hepatocytes or whether, for instance, haematopoietic cells in the liver are defective in the production of one or more factors required for hepatocyte survival. But studies carried out on the regenerating liver have shown that NF- κ B activity is strongly activated during that process, suggesting a direct role for RelA in hepatocyte proliferation¹⁷. □

Received 12 April; accepted 16 May 1995.

- Baeuerle, P. A. & Henkel, T. A. *Rev. Immun.* **12**, 141–179 (1994).
- Sha, W. C., Liou, H.-C., Tuomanen, E. I. & Baltimore, D. *Cell* **80**, 321–330 (1995).
- Wyllie, A. H., Morris, R. G., Smith, A. L. & Dunlop, D. J. *Path.* **142**, 67–77 (1994).
- Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. *J. Cell Biol.* **119**, 493–501 (1992).
- Grilli, M., Chiu, J.-S. & Lenardo, M. J. *Int. Rev. Cytol.* **143**, 1–63 (1991).
- Haskill, S. et al. *Cell* **65**, 1281–1289 (1991).
- Thompson, J. E., Phillips, R. J., Erdjument-Bromage, H., Tempst, P. & Ghosh, S. *Cell* **80**, 573–582 (1995).
- Beg, A. A. & Baldwin, A. S. *Genes Dev.* **7**, 2064–2070 (1993).
- Scott, M. L., Fujita, T., Liou, H.-C., Nolan, G. P. & Baltimore, D. *Genes Dev.* **7**, 1266–1276 (1993).
- Rice, N. R., MacKichan, M. L. & Israel, A. *Cell* **71**, 243–253 (1992).
- Burkly, L. et al. *Nature* **373**, 531–536 (1995).
- Weih, F. et al. *Cell* **80**, 331–340 (1994).
- Abbadie, C. et al. *Cell* **75**, 889–912 (1993).
- Thanos, D. & Maniatis, T. *Cell* **80**, 529–532 (1995).
- Stein, B. et al. *EMBO J.* **10**, 3879–3891 (1993).
- Hilberg, F., Aguzzi, A., Howells, N. & Wagner, E. F. *Nature* **365**, 179–181 (1993).
- Cressman, D. E., Greenbaum, L. E., Haber, B. A. & Taub, R. J. *biol. Chem.* **269**, 30429–30435 (1994).
- Beg, A. A., Finco, T. S., Nantermet, P. V. & Baldwin, A. S. *Molec. cell. Biol.* **13**, 3301–3310 (1993).
- Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
- Nolan, G. P., Ghosh, S., Liou, H.-C., Tempst, P. & Baltimore, D. *Cell* **64**, 961–969 (1991).
- Plump, A. S. et al. *Cell* **71**, 343–353 (1992).
- Liou, H.-C., Sha, W. C., Scott, M. L. & Baltimore, D. *Molec. cell. Biol.* **14**, 5349–5359 (1994).
- Montgomery, R. A. & Dallman, M. J. *J. Immun.* **147**, 554–560 (1991).

ACKNOWLEDGEMENTS. We thank the knockout facility at the Rockefeller University for help in generation of chimaeric mice, and A. Gifford for help in genotyping. We also thank K. Macleod, T. Jacks and members of the Baltimore laboratory for discussion. A.A.B. is supported by a postdoctoral grant from the Cancer Research Institute. W.C.S. is supported by a fellowship from the NIH. This work was supported by the NIH and by the Amgen Corporation.

Control of neuronal pathway selection by a *Drosophila* receptor protein-tyrosine kinase family member

Christopher A. Callahan^{*†}, M. G. Muralidhar^{*},
Scott E. Lundgren^{*}, Audra L. Scully^{*}
& John B. Thomas^{*‡}

^{*} Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186, USA

[†] Department of Neurosciences, University of California at San Diego, La Jolla, California 92093, USA

DURING development, neurons are capable of selecting specific pathways that lead them to their appropriate target areas. A variety of molecular mechanisms are thought to be involved in pathway recognition, including cell adhesion¹, repulsion^{2,3} and chemotropism⁴. However, apart from a few genes whose involvement has been shown genetically⁵⁻⁸, the mechanisms underlying neuronal pathway selection are largely unknown. Here we report the isolation of the *Drosophila* *derailed* (*drl*) gene, which encodes a novel member of the receptor protein-tyrosine kinase family. Using a newly developed axon-targeted reporter gene⁹ we find that *drl* is expressed by a small subset of embryonic interneurons whose growth cones choose common pathways during development. In *drl* mutant embryos these neurons fail to make the correct pathway choices. Our results provide evidence for receptor protein-tyrosine kinase involvement in key aspects of neuronal pathway recognition.

To isolate genes involved in neuronal pathfinding, we carried out a P element enhancer trap screen for genes expressed in restricted subsets of developing neurons that choose common pathways. Unfortunately, the commonly used reporter molecules for enhancer trap vectors¹⁰⁻¹² fail to diffuse readily into axons, and thus the identities and relationships among expressing neurons cannot be determined directly. We therefore incorporated into our vector the *tau-lacZ* reporter gene⁹, whose product is a fusion between the microtubule-associated Tau protein and β -galactosidase (β -gal). This reporter efficiently labels the axonal projections of expressing neurons, presumably owing to the microtubule binding properties of Tau protein. From 2,400 independent insertions of the P[*tau-lacZ*] enhancer trap element, we identified 45 lines in which small subsets of Tau- β -gal-expressing neurons project along common pathways during embryogenesis⁹. The *drl* gene was identified from line 3.765 (hereafter called *drl*^{P3.765}), which contains a single P[*tau-lacZ*] element located at polytene chromosome band 37C and expresses the reporter in a small subset of neurons, muscles and epidermal cells. Here we address the nervous system function of *drl*.

By hour 12 of embryogenesis, there are approximately 200 neurons per hemisegment within the developing ventral nerve cord (VNC), many of which have begun elongating axons. These neurons form an orthogonal framework of axon bundles comprising two bilaterally symmetric longitudinal connectives and an anterior and posterior commissure that span the midline within each segment (Fig. 1a). Within the VNC of *drl*^{P3.765} embryos, Tau- β -gal expression is restricted to a cluster of approximately 20 interneurons per hemisegment (Fig. 1b). Expression commences postmitotically in the neurons as they begin elongating axons, and continues throughout embryogenesis. In embryos heterozygous for *drl*^{P3.765}, these interneurons project in a highly stereotyped fashion. Most if not all of the *drl* neurons extend their axons across the midline within the anterior commissure. On reaching the contralateral longitudinal

connective, their growth cones turn anteriorly and choose to follow one or other of two discrete pathways located close together at the medial edge of the connective. Once the *drl* neurons have reached the adjacent anterior segment, they fasciculate with their homologues, forming two continuous axon fascicles, termed DD and DV, in each connective (Fig. 1c, d).

In *drl*^{P3.765} homozygous mutant embryos, the *drl* neurons cross the midline and reach the contralateral longitudinal connective, but they project anteriorly along inappropriate paths. Instead of choosing their normal medial pathways and fasciculating with one another, *drl* neurons project without any apparent preference for pathways within the connectives and thus fail to form the distinctive DD and DV bundles (Fig. 1e). In many segments the *drl* neurons can be seen crossing one another. *drl* neurons also appear less organized within the anterior commissure, and in approximately 10% of segments *drl* axons can be detected crossing abnormally in the posterior commissure (Fig. 1e, f). These same phenotypes are seen when *drl*^{P3.765} is placed in trans to *drl*^{R343}, a deletion of *drl* created by imprecise excision of the P[*tau-lacZ*] element (Fig. 1f). In contrast, wild-type pathfinding is completely restored when *drl*^{P3.765} is placed in trans to a precise excision of the P element, thus showing that the P element is the cause of the neuronal phenotype (Fig. 1g). Interestingly, both *drl*^{P3.765} and *drl*^{R343} mutant individuals, although viable and fertile, are sluggish and uncoordinated, suggesting that the *drl* pathfinding defects compromise the ability of the *drl* neurons to make proper synaptic connections during development.

Although the *drl* fascicles are dramatically affected in *drl* mutants, the overall structure of the nervous system and the formation of non-*drl* axon bundles, as assayed with antibodies to horseradish peroxidase (HRP)¹³, Fasciclin II¹⁴ or Connectin¹⁵, are indistinguishable from the wild type in both *drl*^{P3.765} and *drl*^{R343} individuals (Fasciclin II staining is shown in Fig. 1h). Thus neuronal pathfinding defects in *drl* mutants are not widespread. To address the further possibility that *drl* mutations may cause general changes in cell fate of the *drl* neurons, we examined the expression of cell-specific neuronal markers^{16,17} in *drl* mutant embryos. From the set of markers examined, which included *apterous*, *engrailed*, *even skipped*, *prospero* and *Sex combs reduced*, only the LIM-homeodomain gene *islet* (S. Thor and J.B.T., unpublished) is expressed by any of the *drl* neurons. In both *drl*^{R343} and *drl*^{P3.765}/*drl*^{R343} embryos, the pattern of *islet* expression is indistinguishable from the wild type (results not shown).

Plasmid-rescued DNA flanking the P[*tau-lacZ*] element in *drl*^{P3.765} was used to isolate genomic and complementary DNA clones. We found that the *drl* gene maps ~70 kilobases (kb) proximal to the *Dopa decarboxylase* (*Ddc*) gene¹⁸, but does not correspond to any of the known complementation groups in the region^{19,20} (Fig. 2). The P[*tau-lacZ*] element is inserted within the *drl* gene, 47 base pairs (bp) upstream from the 5' end of the longest *drl* cDNA. The cell body pattern of *drl* RNA expression in the wild type is identical to that of Tau- β -gal in *drl*^{P3.765} embryos (Fig. 3a), indicating that the Tau- β -gal reporter expression is indeed reflecting the expression of the *drl* gene. In *drl*^{P3.765} homozygous embryos, *drl* transcript levels within the nervous system are almost completely eliminated and are detectable only after overdevelopment for colour reaction (Fig. 3b). Wild-type levels and patterns of *drl* expression are restored by precise excision of the P element, proving that the P element insertion causes the disruption of expression in *drl*^{P3.765} embryos.

The longest *drl* cDNA contains a single long open reading frame (ORF) that encodes a new member of the receptor protein-tyrosine kinase (RPTK) family. RPTKs comprise a well-studied family of transmembrane receptors with intrinsic protein-tyrosine kinase activity and are capable of activating intracellular signalling pathways on ligand binding²¹. RPTKs have well documented roles in a variety of developmental events, including the control of cell fate^{22,23}, cell migration²⁴, and neu-

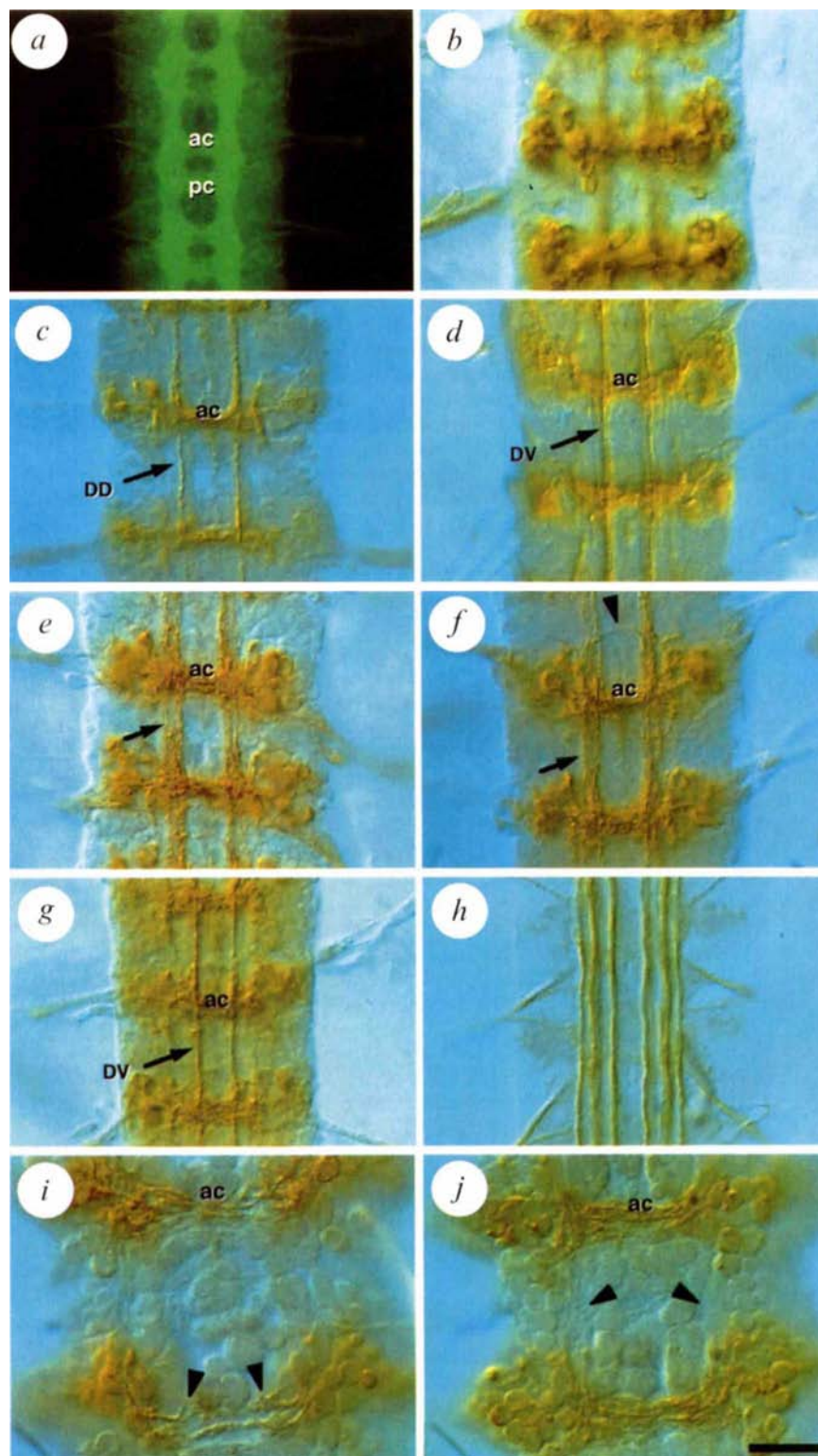
[‡] To whom correspondence should be addressed.

ronal outgrowth and survival *in vitro*²⁵. Drl shows the highest similarity to mammalian RYK^{26, 28} (Fig. 4). Both Drl and RYK have novel, short extracellular domains that contain distinct regions of amino-acid identity. The two proteins also share unique amino-acid substitutions within the catalytic domain, which, together with their extracellular domains, place them in a new RPTK subfamily. Despite the catalytic domain substitutions, the presence of those amino acids thought to be essential for protein-tyrosine kinase function predicts that both Drl and RYK are likely to be catalytically active²¹.

Using an antibody raised against Drl fusion protein, we found that Drl is localized to the growth cones and axons of the *drl* neurons as they cross within the anterior commissure (Fig. 1i). Interestingly, after the neurons have turned anteriorly, Drl cannot be detected on the portion of the axons within the connectives, and remains restricted to the axon segments within the anterior commissure (Fig. 1j). This suggests that Drl does not physically participate in the fasciculation of *drl* neurons within the DD and DV bundles. Instead, our results suggest the existence of multiple, discrete steps in neuronal pathway recognition,

FIG. 1 *drl* mutants have neuronal pathfinding defects. a, Fluorescein-isothiocyanate-conjugated anti-horseradish peroxidase staining of a wild-type embryo. Neurons form a highly regular framework of axon bundles within the VNC, consisting of the bilaterally symmetric longitudinal connectives and, in each segment, an anterior and posterior commissure (ac and pc, respectively). b–g, Embryos stained with antibodies to β -gal to visualize Tau- β -gal expression in the *drl* neurons. b, An hour-12 *drl*^{P3.765} embryo showing expression in approximately 20 neurons per hemisegment. The focal plane is at the level of the neuronal cell bodies. The *drl* axon bundles within the longitudinal connectives are out of the focal plane. c, d, *drl*^{P3.765/+} embryos. The *drl* neurons are interneurons that project across the midline within the anterior commissure and extend axons medially and anteriorly within the contralateral connectives. By hour 11.5, the *drl* neurons have fasciculated with their homologues in adjacent anterior segments, forming the DD and DV axon bundles. e, f, A *drl*^{P3.765} (e) and a *drl*^{P3.765/drl}^{R343} (f) embryo. Both exhibit the same pathfinding defects. The *drl* neurons project across the midline where the axons often appear less tightly fasciculated than in the wild type; the arrowhead in (f) points to an axon abnormally crossing the midline in the posterior commissure. Within the contralateral connectives the *drl* neurons fail to choose their appropriate medial pathways, and instead project along many inappropriate paths, including those in very lateral positions (arrows). g, A *drl*^{P3.765/R193} embryo shows normal pathfinding which results in normal *drl* bundles. R193 is a precise excision of the *drl*^{P3.765} P element as judged by Southern analysis. h, A *drl*^{R343} embryo stained with a monoclonal antibody to Fasciclin II¹⁴. The Fasciclin II axon bundles are indistinguishable from wild-type. i, j, An hour-10 (i) and hour-11.5 (j) embryo stained with anti-Drl antibody. At hour-10, staining is detected on the cell bodies, axons and the growth cones (arrowheads in i) of the *drl* neurons as they extend within the anterior commissure. At hour 11.5, although many of the *drl* neurons have already turned anteriorly within the connective (arrowheads in j) and reached the adjacent anterior segment (compare with Tau- β -gal-stained embryos in c, d), staining is detected only on the portion of the *drl* axons within the anterior commissure. Anterior is up. Scale bar, 30 μ m for a, 20 μ m for b–j.

METHODS. Rat polyclonal sera were raised against a glutathione S-transferase fusion protein containing amino acids 19–234 of the Drl extracellular domain; antibody was pre-absorbed with *drl*^{R343} embryos before use. Staining can be detected in live embryos, indicating that Drl is localized to the membrane. Embryos were dissected, fixed and stained as described⁹.



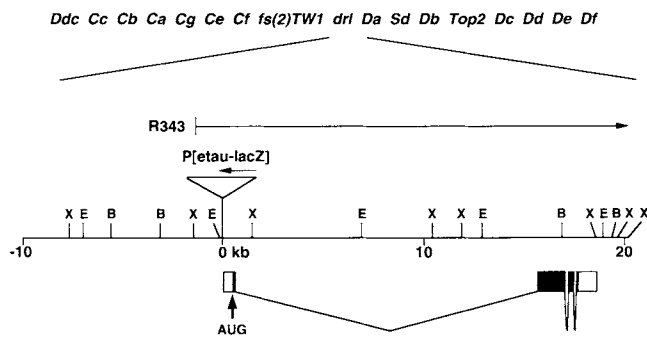


FIG. 2 Genomic organization of the *drl* gene. At the top is a map of the complementation groups in the *drl* region^{19,20}. *drl* lies between *fs(2)TW1* and *I(2)37Da* and is located 70 kb to the right of *Ddc*. Below the genetic map is the molecular map of the *drl* gene. Restriction sites shown are *Bam*HI (B), *Eco*RI (E) and *Xho*I (X). *drl* is spliced from 4 exons distributed over 18 kb. The closed boxes correspond to the *drl* ORF. The P[etau-lacZ] P element is located 47 bp upstream of the 5' end of the longest *drl* cDNA and 395 bp upstream from the *drl* ORF. The P element is shown 1/5 scale, and arrow denotes orientation of tau-lacZ transcription from the P promoter. Directly above the restriction map is the extent of the R343 excision-induced deletion that removes the entire *drl* ORF.

METHODS. A 3.5-kb *Bgl*II genomic fragment flanking the P element was plasmid-rescued from *drl*^{P3.765} DNA and used to isolate genomic and *drl* cDNA clones. Four overlapping cosmid clones spanning 60 kb were isolated and used to link *drl* to a previous genomic walk in the 37B-C region¹⁸. The longest *drl* cDNA (3.1 kb) was sequenced. The intron-exon structure of *drl* and the location of the *drl*^{P3.765} P-element integration site were determined by sequencing genomic DNA.

FIG. 3 Expression of *drl* is disrupted in *drl*^{P3.765} mutants. Ventral views of wild-type (a) and *drl*^{P3.765} (b) whole-mount embryos hybridized with a digoxigenin-labelled *drl* antisense riboprobe. The *drl*^{P3.765} embryo was over-developed for alkaline phosphatase colour reaction in comparison to the wild-type embryo. In wild-type embryos, the cell body expression pattern is indistinguishable from that of Tau-β-gal from the *drl* P[etau-lacZ] element. In contrast, transcript levels within the nervous system of *drl*^{P3.765} are almost entirely eliminated. Expression is detectable only in a few of the *drl* neurons and in these only at extremely low levels (arrows compare homologous neuronal clusters in wild-type and *drl*^{P3.765} embryos). Anterior is up. Scale bar, 20 μm.

METHODS. Whole-mount *in situ* hybridizations were done as described^{17,29}. Digoxigenin-labelled sense and antisense riboprobes of *drl* were prepared from the 3.1-kb *drl* cDNA. After staining, embryos were mounted directly in 90% glycerol or dehydrated in an ethanol series, briefly rinsed in xylenes and mounted in Permount.

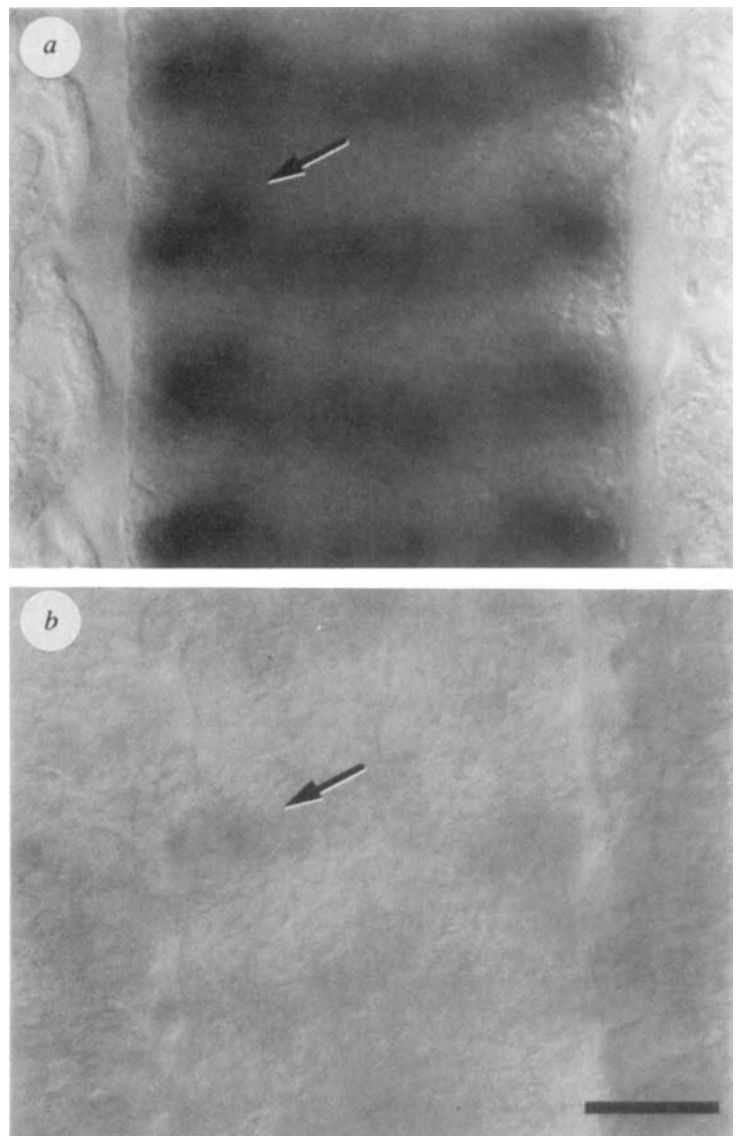
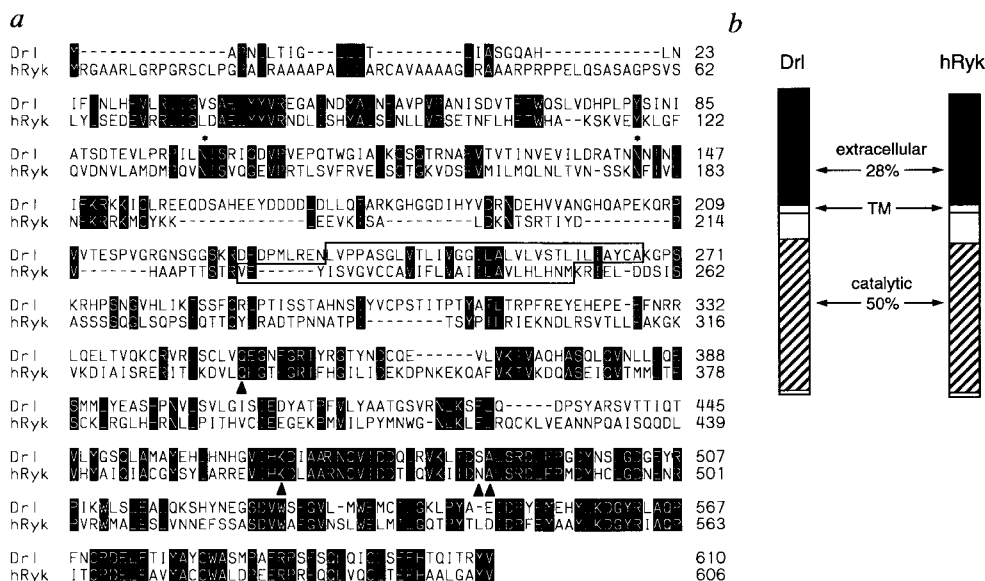


FIG. 4 *drl* encodes an RPTK related to RYK. **a**, Deduced amino-acid sequence of Drl aligned with human RYK (hRYK)²⁸. Identities are shown white on black. Unusual amino-acid substitutions shared between the catalytic domains of Drl and hRYK are denoted by arrowheads. The catalytic loop motif HRDLAARN is HKDLAARN in Drl and HKDLAARN in hRYK. The highly conserved DFG motif is DSA in Drl and DNA in hRYK, and the first glycine within the GXGXXG motif of both proteins is instead a glutamine. As noted for RYK²⁶, the substitutions in these latter two motifs may be compensatory because the two regions lie in close proximity to one another³⁰ and probably cooperate in binding ATP. Asterisks denote possible N-linked glycosylation sites conserved between Drl and hRYK. The putative transmembrane domains of both proteins are boxed. **b**, Comparison of the topology of Drl and hRYK and the percentage identities between their extracellular and catalytic domains.



with Drl playing a key controlling role. We predict that once activated by its ligand, Drl controls pathway selection by modulating the function of specific cell adhesion molecules that mediate proper axon fasciculation. These results in *Drosophila* provide direct *in vivo* evidence for RPTKs' playing key roles in neuronal pathfinding and axon fasciculation. □

Received 30 May; accepted 8 June 1995.

1. Rutishauser, U. *Curr. Opin. Neurobiol.* **3**, 709–715 (1993).
2. Luo, Y., Raible, D. & Raper, J. A. *Cell* **75**, 217–227 (1993).
3. Nose, A., Takeichi, M. & Goodman, C. S. *Neuron* **13**, 525–539 (1994).
4. Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. *Cell* **78**, 425–435 (1994).
5. Elkins, T., Zinn, K., McAllister, L., Hoffmann, F. M. & Goodman, C. S. *Cell* **60**, 565–575 (1990).
6. Leung-Hagstelt, C. et al. *Cell* **71**, 289–299 (1992).
7. Lin, D. M., Fetter, R. D., Kocyanski, C., Grenningloh, G. & Goodman, C. S. *Neuron* **13**, 1055–1069 (1994).
8. Thomas, J. B. & Wyman, R. J. *Nature* **298**, 650–651 (1982).
9. Callahan, C. A. & Thomas, J. B. *Proc. natn. Acad. Sci. U.S.A.* **91**, 5972–5976 (1994).
10. O'Kane, C. J. & Gehring, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9123–9127 (1987).
11. Smith, H. K. & O'Kane, C. J. *Wilhelm Roux Arch. dev. Biol.* **200**, 306–311 (1991).
12. Bier, E. et al. *Genes Dev.* **3**, 1273–1287 (1989).
13. Jan, L. Y. & Jan, Y. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2700–2704 (1982).
14. Van Vactor, D., Sink, H., Fambrough, D., Tsou, R. & Goodman, C. S. *Cell* **73**, 1155–1164 (1993).
15. Nose, A., Mahajan, V. B. & Goodman, C. S. *Cell* **70**, 553–567 (1992).
16. Doe, C. Q. & Goodman, C. S. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez Arias, A.) 1131–1206 (Cold Spring Harbor Lab., Cold Spring Harbor, 1993).
17. Bourgouin, C., Lundgren, S. E. & Thomas, J. B. *Neuron* **9**, 549–561 (1992).
18. Gilbert, D., Hirsh, J. & Wright, T. R. F. *Genetics* **106**, 679–694 (1984).
19. Wright, T. R. F. et al. *Chromosoma* **83**, 45–58 (1981).
20. Gay, P. & Contamine, D. *Molec. gen. Genet.* **239**, 361–370 (1993).
21. van der Geer, P., Hunter, T. & Lindberg, R. A. *Rev. Cell Biol.* **10**, 251–337 (1994).
22. Hafez, E., Basler, K., Edstrom, J.-E. & Rubin, G. M. *Science* **236**, 55–63 (1987).
23. Aroian, R. V., Koga, M., Mendel, J. E., Ohshima, Y. & Sternberg, P. W. *Nature* **348**, 693–698 (1990).
24. Klamt, C., Glazer, L. & Shilo, B.-Z. *Genes Dev.* **6**, 1668–1678 (1992).
25. Chao, M. V. *Neuron* **9**, 583–593 (1992).
26. Hovens, C. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11818–11822 (1992).
27. Maminta, M. L. D. et al. *Biochem. biophys. Res. Commun.* **189**, 1077–1083 (1992).
28. Stacker, S. A. et al. *Oncogene* **8**, 1347–1356 (1993).
29. Jiang, J., Kosman, D., Ip, Y. T. & Levine, M. *Genes Dev.* **5**, 1881–1891 (1991).
30. Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. *Nature* **372**, 746–754 (1994).

ACKNOWLEDGEMENTS. We thank D. Stathakis and T. Wright for supplying their genomic walk and sharing valuable insights; S. Thor and C. Q. Doe for supplying probes; D. Contamine for supplying stocks; J. Nolan, D. Wettstein, K. Scott and W.-C. Jen for assistance; T. Hunter, G. Lemke, M. Goulding, A. Bang and S. Thor for critical reading of the manuscript; and A. Wilks and S. Stacker for discussions. This work was supported by grants from the NIH, a March of Dimes Basil O'Connor Scholar Research Award and a Pew Scholars Award from the Pew Memorial Trusts to J.B.T., and from a NIH/NIGMS Training Grant to C.A.C., who is a student in the Medical Scientist Training Program at UCSD.

Hippocampal GABA transporter function in temporal-lobe epilepsy

Matthew J. During, Kathryn M. Ryder & Dennis D. Spencer

Molecular Pharmacology and Neurogenetics Laboratory, Sections of Neurosurgery and Endocrinology, Departments of Surgery and Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

ELECTROPHYSIOLOGICAL studies of human temporal-lobe epilepsy suggest that a loss of hippocampal GABA-mediated inhibition may underlie the neuronal hyperexcitability^{1–3}. However, GABA (γ-aminobutyric acid)-containing cells are preserved⁴ and GABA receptors are maintained in the surviving hippocampal neurons⁵. Diminished GABA release may therefore mediate the loss of inhibition. Here we show that, in the human brain, potassium-stimulated release of GABA was increased, and glutamate-induced, calcium-independent release of GABA was markedly decreased, in epileptogenic hippocampi, in contrast with contralateral, non-epileptogenic hippocampi. The glutamate-induced GABA release *in vivo* was transporter-mediated in rats. Furthermore, in amygdala-kindled rats, a model for human epilepsy, a decrease in glutamate-induced GABA release was associated with a 48% decrease in the number of GABA transporters. These data suggest that temporal-lobe epilepsy is characterized in part by a loss of glutamate-stimulated GABA release that is secondary to a reduction in the number of GABA transporters.

A human microdialysis probe enabling the direct measurement of neurotransmitter release within specific regions of the conscious human brain has been developed and characterized⁶. This technique was used to test the hypothesis directly that the epileptogenic region in temporal-lobe epilepsy is characterized by an alteration in the regulation of GABA release.

Patients with complex partial seizures refractory to antiepileptic drugs were evaluated with depth electrodes. Bilateral hippocampal probes were stereotactically implanted by using MRI-directed coordinates, and epileptogenic hippocampi were identified by monitoring of spontaneous seizures⁶. Patients gave fully informed consent to a protocol approved by the Yale University Human Investigation Committee.

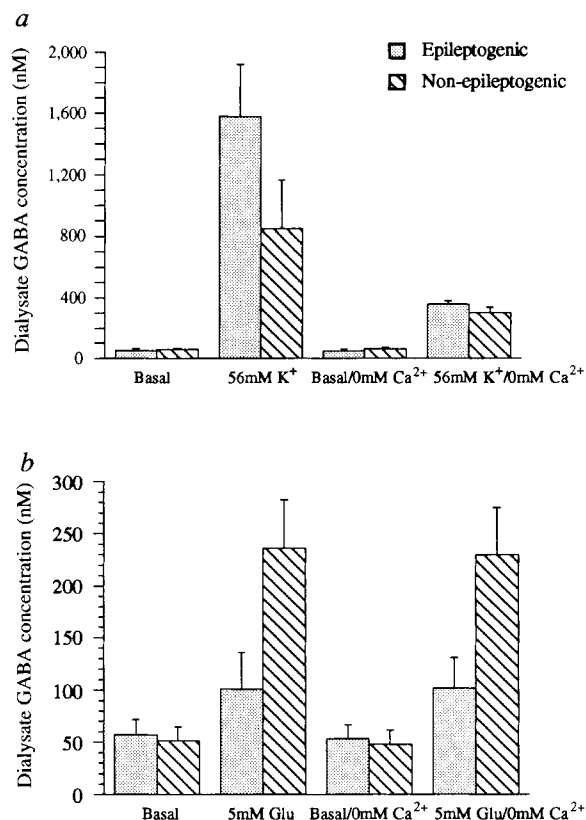


FIG. 1 GABA release in the human hippocampus. *a*, Concentrations of GABA were increased from baseline in both epileptogenic and non-epileptogenic hippocampi ($P < 0.01$); however, the potassium-stimulated concentration was greater in the epileptogenic hippocampus ($P < 0.05$; two-way repeated-measures analysis of variance (ANOVA) with Fisher post-hoc test). *b*, The stimulated release in both hippocampi was dependent on extracellular calcium ($P < 0.01$). In the non-epileptogenic hippocampus there was a significant increase in glutamate-induced GABA release ($P < 0.01$, repeated-measures ANOVA with Fisher post-hoc test). In contrast, there was no effect of glutamate on GABA concentrations in the epileptogenic hippocampi. Moreover, the glutamate-induced release in the normal hippocampus was independent of extracellular calcium.

METHODS. Eight patients underwent microdialysis while awake, as previously described⁶. After a 90-min baseline under calcium-free or physiological (1.2 mM) Ca²⁺ perfusion, the perfusion fluid was changed to include 56 mM K⁺. After a minimum of 6 h the stimulation was repeated under the alternative (with or without calcium) condition. At least 8 h after the potassium stimulation, six of the original eight patients underwent microdialysis with or without calcium in the perfusion fluid. After collection of a 30-min baseline sample, the perfusion fluid was changed to include 5 mM glutamate. The glutamate challenge was repeated under alternative calcium conditions at least 6 h apart. This dose of glutamate was chosen after pilot studies in both rats and humans to determine the minimum concentration to obtain a consistent effect on GABA release. Although extremely high concentrations of glutamate pressure-injected intraparenchymally into the brain may cause excitotoxicity and neuronal injury, the microdialysis system could deliver a maximum of 6.25 nmol of glutamate per min assuming 100% extraction of the glutamate into the brain. As the efficiency of the membrane is approximately 30% the actual delivery is approximately 2 nmol min⁻¹. We (unpublished results) and others²⁶ have shown that glutamate perfusion at concentrations three orders of magnitude greater is completely safe and non-toxic in studies involving experimental animals.

TABLE 1 ³H-Nipecotic acid binding in homogenates of rat hippocampus

Treatment	K _d (μM)	B _{max} (nmol per mg protein)
Control	1.90 ± 0.22	211 ± 42
Kindled	1.85 ± 0.44	110 ± 20*

Animals implanted with amygdala electrodes were kindled to stage 5 ($n = 6$) or handled ($n = 6$) as previously reported²⁵. Kindled rats and paired surgical controls were used at least one week after the last seizure. Animals were decapitated under CO₂ anaesthesia, and hippocampi were rapidly dissected and frozen in solid CO₂ and stored at -80 °C. For analysis, tissue was quickly thawed, homogenized in 50 volumes of 50 mM Tris acetate buffer, pH 7.4, containing 400 mM NaCl and then centrifuged at 15,000g for 10 minutes at 4 °C. The pellet was washed once by rehomogenization and centrifugation. The resulting pellet was suspended and sonicated in 50 volumes of 50 mM Tris acetate buffer, pH 7.4, containing 400 mM NaCl. Portions (160 μl) of the crude membrane preparation were added to 40 μl of labelled nipecotic acid (25.8 Ci mmol⁻¹, Amersham) for a final concentration of 100 nM. Samples were incubated at 4 °C for 90 min, and incubation was terminated by rapid filtration through Whatman GF/C filters with a Brandel cell harvester. Filters were washed immediately with ice-cold Tris acetate buffer, pH 7.4. Radioactivity was assayed with a scintillation counter. Non-specific binding, defined by the addition of 20 mM GABA, was less than 10%. Five concentrations of cold nipecotic acid ranging from 300 to 9,900 nM were used to displace the tritiated ligand. Assays were run in triplicate. Scatchard plots were analysed with the EBDA program, Kell 2.0 (Elsevier, Amsterdam), to determine the equilibrium dissociation constant (K_d) and the total number of ³H-nipecotic acid binding sites (B_{max}).

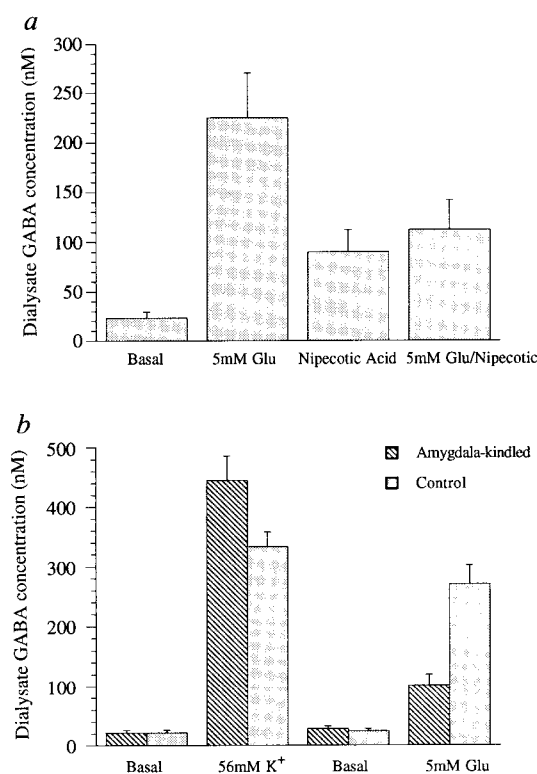
* $P < 0.05$, Student's *t*-test.

On days 7–16 after implantation of the microdialysis/depth electrodes and a minimum of 24 hours after the last seizure, patients underwent microdialysis perfusion. The protocol included three phases to assess: first, potassium-depolarization-induced release of GABA and its dependency on extracellular calcium; secondly, glutamate-induced GABA release and calcium dependency; and thirdly, the extraction efficiency of GABA from dialysate into hippocampal extracellular fluid.

Basal GABA concentrations were similar in dialysate from epileptogenic hippocampi and the homotopic sites of the contralateral hemisphere: 50 ± 14 nM and 59 ± 8 nM, respectively. However, potassium-stimulated, calcium-dependent release was significantly higher in the epileptogenic hippocampus (32 ± 7 -fold and 14 ± 5 -fold increases, respectively). In contrast, the local perfusion of glutamate induced a 4.6 ± 0.9 -fold calcium-independent increase in GABA concentrations in the non-epileptogenic hippocampus compared with a non-significant 1.7 ± 0.6 -fold increase in the epileptogenic hippocampus (Fig. 1).

The increased GABA concentration in the dialysate obtained during potassium-induced stimulation might reflect either enhanced release or diminished re-uptake. GABA re-uptake occurs via both neuronal and glial transporters⁷; however, these transporters are electrogenic and, after cellular depolarization associated with a sustained evaluation in intracellular sodium concentration, may function in reverse, thereby increasing extracellular levels of GABA⁸. Glutamate-induced release of GABA has been shown to be blocked by nipecotic acid, a GABA transport inhibitor, suggesting transporter reversal as the mechanism *in vitro*⁹. In contrast, potassium-induced release of GABA is enhanced in the presence of nipecotic acid¹⁰. To determine whether transporter reversal mediated the glutamate-induced GABA release in the human hippocampus *in vivo*, we studied the effect of nipecotic acid in the rat. Nipecotic acid completely antagonized the glutamate-induced rise in extracellular GABA (Fig. 2). We also studied amygdala-kindled rats, an established model for human temporal-lobe epilepsy¹¹. The potassium-induced release was potentiated, whereas the glutamate-induced release was markedly attenuated (Fig. 2). These results were therefore in concordance with the data from humans. To deter-

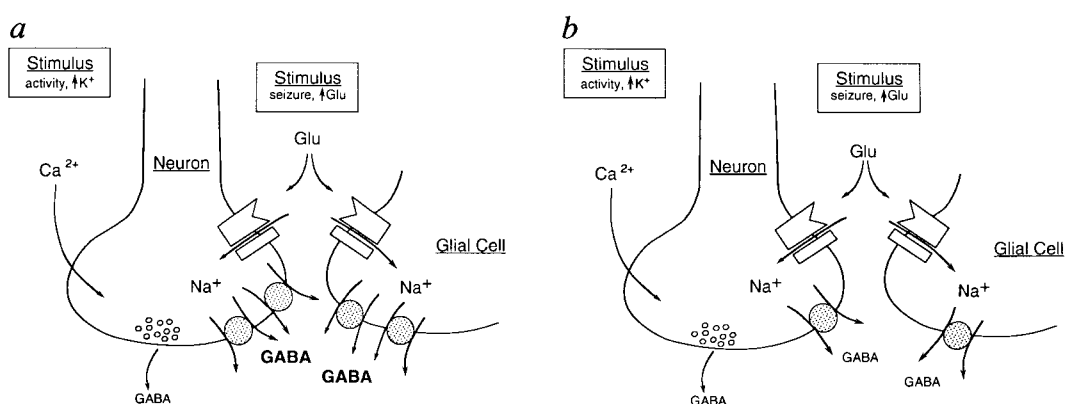
mine whether the loss of glutamate-induced response reflected a loss of GABA transporters, binding of ^3H -nipecotic acid¹² was carried out in hippocampus obtained from both naive, control animals and amygdala-kindled rats. The K_d of ^3H -nipecotic acid binding was unchanged, whereas there was a $48 \pm 9\%$ reduction in B_{max} in amygdala-kindled rats (Table 1). A further microdialysis study was carried out to assess human GABA transporter function *in vivo*. We used a reverse dialysis method¹³ to determine the extraction of exogenous GABA from dialysis fluid into hippocampal extracellular fluid. We established the validity of this method by performing studies in rats by using local perfusion of a potent GABA transporter inhibitor, tiagabine¹⁴. At a tiagabine concentration of $10 \mu\text{M}$, the extraction of GABA from the perfusate was diminished by $21 \pm 5\%$. In humans, local perfusion of GABA ($10 \mu\text{M}$) resulted in a $32 \pm 4\%$ extraction in the epileptogenic hippocampus, whereas the extraction in the non-epileptogenic hippocampus was $39 \pm 4\%$ (mean $\pm$ s.e.m., $P < 0.05$, *t*-test).



Glutamate induces GABA release by depolarization of neurons by binding to ligand gated channels, both NMDA (*N*-methyl-D-aspartate) receptors and non-NMDA receptors¹⁵. Glutamate may also stimulate GABA release from glia through non-NMDA receptors¹⁶. A loss of either of these receptor classes in either GABA neurons or glia would also be consistent with our data. However, both NMDA and non-NMDA electrophysiological responses are maintained, and possibly enhanced, in resected epileptogenic tissue^{1,3}. Furthermore, receptor autoradiography studies have generally reported a relative preservation of glutamate receptors or an increase in ligand binding¹⁷⁻¹⁹. These data therefore suggest that the loss of glutamate-induced GABA release is not secondary to a loss of glutamate receptors but is more consistent with an alteration of the release machinery of GABA. Moreover, two other studies support our hypothesis that GABA transporter function is abnormal in epileptogenic tissue. First, following the acute oral administration of tiagabine in patients with temporal-lobe epilepsy there is a diminished

FIG. 2 GABA release in the rat hippocampus. *a*, 5 mM glutamate induced a significant increase in GABA release ($P < 0.05$) in the normal rat ($n = 5$), but nipecotic acid, which raised basal GABA release ($P < 0.05$), blocked the glutamate-stimulated release ($P < 0.05$; two-way repeated-measures ANOVA with Fisher post-hoc test). *b*, Potassium (56 mM) induced an increase in hippocampal GABA release in both control ($n = 6$) and kindled ($n = 6$) rats ($P < 0.01$), with higher concentrations reached in the kindled animals ($P < 0.05$). The glutamate-induced rise in GABA was significant in both control ($P < 0.01$) and amygdala-kindled rats ($P < 0.05$), although the increase was greater in the control animals ($P < 0.05$; repeated-measures ANOVA with Fisher post-hoc test). METHODS. Adult male Sprague-Dawley rats were anaesthetized and a guide cannula was implanted above the hippocampus (-4.3 anteroposterior, -3.0 lateral). Dialysis probes were implanted one week after surgery, with the exposed membrane spanning the dorsal hippocampus. At 20 h after implantation, awake, freely-moving animals were perfused with artificial extracellular fluid (aECF) as previously described²⁵. Following a 2-h stabilization period, a baseline sample was taken over 20 min. The dialysis fluid was changed to include glutamate (5 mM) with a further 20-min sample collected. The perfusion fluid was changed back to regular artificial ECF for 2.5 h and then nipecotic acid (1 mM) was perfused for 1 h with a baseline taken during the final 20 min. Glutamate (5 mM) was then added to the nipecotic acid-containing ECF and a further 20-min sample taken. Repeated glutamate perfusion in the absence of nipecotic acid showed no attenuation in GABA release at the same stimulus interval. A second group of adult male Sprague-Dawley rats were amygdala-kindled as previously described²⁵. At 20 h after implantation, awake, freely-moving animals underwent microdialysis perfusion. Following a 2-h stabilization period, a baseline sample was taken over 20 min. The dialysis fluid was changed to include glutamate (5 mM) and a further 20-min sample was collected.

FIG. 3 A schematic diagram showing release of GABA from both neurons and glia after two classes of stimulation: high K⁺ and normal synaptic activity, and elevated glutamate and seizure activity. *a*, Normal hippocampus. After high-potassium-induced stimulation or an increase in neuronal firing, calcium enters the nerve terminal and GABA is released extracellularly from vesicles via classical, quantal exocytosis. With an increase in extracellular glutamate concentration, sodium enters through ionotropic glutamate receptors and drives the electrogenic GABA transporter in the outward direction, resulting in GABA release from both neurons and glia. *b*, Epileptogenic hippocampus, in which the high-potassium-induced vesicular release is maintained. However,



in response to elevated glutamate concentration, as may be seen with excessive excitation before seizure generation⁶, there is a decrease in the generation of extracellular GABA, reflecting the loss of GABA transporters in neurons and/or glia.

effect on increasing extracellular GABA in epileptogenic hippocampi²⁰. Second, GABA-induced responses in granule cells recorded from epileptogenic hippocampal slices are markedly prolonged²¹. The channel characteristics are unchanged and this response is mimicked in control tissue treated with a GABA transporter inhibitor, whereas these transporter inhibitors have minimal effect on sclerotic hippocampi.

In summary, our data demonstrate an alteration in stimulus-induced release of GABA in human epileptic tissue *in vivo*. We have previously shown that the seizure-induced increase in extracellular GABA is diminished in epileptogenic hippocampi⁶. As potassium-induced release is intact, these data suggest that the release of GABA during seizures may in part be mediated via a calcium-independent, transporter reversal mechanism. Indeed, in one patient in whom a spontaneous seizure occurred during perfusion of calcium-free media, extracellular GABA increased to concentrations within the range previously reported⁶. Transporter function, although critical for terminating fast synaptic transmission, may also be essential for the maintenance of normal inhibition. Several studies have shown a paradoxical effect of GABA transporter inhibitors to facilitate seizures^{22,23}. Although this response might reflect GABA acting extrasynaptically, particularly on dendrites, where GABA produces depolarizing responses²⁴, this result is also consistent with transporter-mediated inhibitory GABA-mediated transmission. We therefore propose that during periods of intense neuronal and glial excitation, release of the inhibitory transmitter is mediated in part by non-vesicular mechanisms and may provide a mechanism for 'dumping' GABA extracellularly and diffusely increasing inhibitory tone (Fig. 3). This calcium-independent release of GABA may be critical to maintain normal inhibition during periods of intense excitation and elevated extracellular glutamate concentration, which occurs before seizure generation⁶. The loss of GABA transporters would diminish this alternative release mechanism limiting inhibition and contributing to the epileptogenic state in human temporal-lobe epilepsy. □

Received 9 January; accepted 11 May 1995.

- Isokawa, M., Avanzini, G., Finch, D. M., Babb, T. L. & Levesque, M. F. *Epilepsy Res.* **9**, 242–250 (1991).
- Knowles, W. D., Awad, I. A. & Nayel, M. H. *Epilepsia* **33**, 601–609 (1992).
- Williamson, A. Clin. *Neurosci.* **2**, 47–52 (1994).
- Babb, T. L., Pretorius, J. K., Kufer, W. R. & Crandall, P. H. *J. Neurosci.* **9**, 2562–2574 (1989).
- Johnson, E. et al. *Neurology* **42**, 811–815 (1992).
- During, M. J. & Spencer, D. D. *Lancet* **341**, 1607–1610 (1993).
- Radian, R., Ottersen, O. P., Storm-Mathisen, J., Castel, M. & Kanner, B. I. *J. Neurosci.* **10**, 1319–1330 (1990).
- Pin, J.-P. & Bockaert, J. *J. Neurosci.* **9**, 648–656 (1989).
- Harris, K. M. & Miller, R. J. *Brain Res.* **482**, 23–33 (1989).
- Szerb, J. C. *J. Neurochem.* **39**, 850–858 (1982).
- McNamara, J. O., Byrne, M. C., Dasheiff, R. M. & Fitz, J. G. *Prog. Neurobiol.* **15**, 139–159 (1980).
- Simpson, M. D., Slater, P., Deakin, J. F., Royston, M. C. & Sken, W. J. *J. Neurosci. Lett.* **107**, 211–215 (1988).
- Morrison, P. F. et al. in *Microdialysis in the Neurosciences* (eds Robinson, T. E. & Justice, J. B.) 47–80 (Elsevier, Amsterdam, 1991).
- Braestrup, C. et al. *J. Neurochem.* **54**, 639–646 (1990).
- Drejer, J., Honore, T. & Schousboe, A. *J. Neurosci.* **7**, 2910–2916 (1987).
- Gallo, V., Patrizio, M. & Levi, G. *Glia* **4**, 245–255 (1991).
- Geddes, J. W. et al. *Exp. Neurol.* **108**, 214–220 (1990).
- McDonald, J. W. et al. *Ann. Neurol.* **29**, 325–332 (1991).
- de Lanerolle, N. C., Brines, M., Williamson, A., Kim, J. H. & Spencer, D. D. in *The Dentate Gyrus and Its Role in Seizures* (eds Mathisen, J., Castel, M. & Moody, I.) 235–250 (Elsevier, Amsterdam, 1991).
- During, M. J. et al. *Epilepsia* **33**, Suppl. 3, 83 (1992).
- Williamson, A., Telford, A. & Spencer, D. D. *J. Neurophysiol.* (in the press).
- Horton, R. W., Collins, J. F., Anlezark, G. M. & Meldrum, B. S. *Eur. J. Pharmac.* **59**, 75–83 (1979).
- Gonsalves, S. F., Twitchell, B., Harbough, R. E., Krosgaard-Larsen, P. & Schousboe, A. *Epilepsy Res.* **4**, 34–41 (1989).
- Alger, B. E. & Nicoll, R. A. *J. Physiol.* **328**, 125–141 (1982).
- Mangano, R. M. & Schwarcz, R. *Brain Res. Bull.* **10**, 47–51 (1983).
- During, M. J., Craig, J. S., Hernandez, T. O., Anderson, G. M. & Gallagher, D. W. *Brain Res.* **584**, 36–44 (1992).

ACKNOWLEDGEMENTS. We thank the medical and nursing staff of Yale–New Haven Hospital for patient evaluation and care. We thank P. DeCamilli, B. Ransom and A. Williamson for helpful discussions and manuscript review. This work was supported in part by grants from the NIH and the Johnson & Johnson Focused Giving Program to M.J.D.

Peripheral deletion of antigen-reactive T cells in oral tolerance

Yuhai Chen, Jun-ichi Inobe, Reinhard Marks, Patricia Gonnella, Vijay K. Kuchroo & Howard L. Weiner*

Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, LMRC 102a, Boston, Massachusetts 02115, USA

ORAL administration of antigen is used to induce antigen-specific peripheral immune tolerance^{1,2}. As well as preventing systemic immune responses to ingested proteins³, oral tolerance to autoantigens has also been used to suppress autoimmune diseases in animals^{4–10} and humans^{11,12}. Both active suppression and clonal anergy are suggested to be mechanisms of oral tolerance, depending on the dose of antigen fed^{13,14}. Here we report that oral antigen can delete antigen-reactive T cells in Peyer's patches, in mice transgenic for the ovalbumin-specific T-cell receptor genes. The deletion was mediated by apoptosis, and was dependent on dosage and frequency of feeding. At lower doses deletion was not observed; instead there was induction of antigen-specific cells that produced transforming growth factor (TGF)- β and interleukin (IL)-4 and IL-10 cytokines. At higher doses, both Th1 and Th2 cells were deleted following their initial activation, whereas cells which secrete TGF- β were resistant to deletion. These findings demonstrate that orally administered antigen can induce tolerance not only by active suppression and clonal anergy but by extrathymic deletion of antigen-reactive Th1 and Th2 cells.

Mice transgenic for the ovalbumin-specific T-cell receptor (TCR) genes which express the V α 13/V β 8.2 TCR on 97% of peripheral T cells¹⁵ were fed up to five times with increasing doses of ovalbumin (0.5 to 500 mg). Depending on the dosage and frequency of feeding, there was either an increase or a decrease in the number of CD4⁺, V β 8.2⁺ T cells in the Peyer's patches (Fig. 1). In animals given 500 mg ovalbumin the percentage of CD4⁺, V β 8.2⁺ T cells had decreased by the time of the third feeding from 20% (unfed) to less than 1.5%. A decrease was also seen in animals given 5 mg ovalbumin, and feeding 50 mg was similar to 5 mg feeding (not shown). The decrease was not the result of an increase in the population of non-T cells as feeding was also associated with a 10 to 25% decrease in the total number of Peyer's patch cells. However, in animals fed 0.5 mg ovalbumin there was a progressive increase in the percentage of the T cells which reached 40% by the fifth feeding. Furthermore, in mice fed 500 mg ovalbumin five times there was a 10 to 20% reduction of CD4⁺, V β 8.2⁺ T cells in the spleen, thymus and all other lymphoid tissues examined. No infiltration of CD4⁺, TCR V β 8.2⁺ T cells was observed in the liver or lung after 500 mg ovalbumin had been given three to five times, demonstrating that T cells do not migrate to other lymphoid or non-lymphoid organs to undergo apoptosis, as has been observed in some systems¹⁶. No effect was observed when mice transgenic for the ovalbumin-specific TCR genes were fed 500 mg of bovine serum albumin, and feeding with 500 mg ovalbumin had no effect on the T-cell frequency in Peyer's patches of non-transgenic animals.

To determine whether the loss of CD4⁺, V β 8.2⁺ cells occurred by deletion, we measured the percentage of apoptotic cells (Fig. 2). In animals given 500 mg ovalbumin there was a marked increase in apoptotic cells following the second feeding which returned to background levels by the third feeding. In the animals given 5 mg, there was a progressive increase in apoptotic

* To whom correspondence should be addressed.

cells which reached 8% by the fifth feeding, whereas only a few cells underwent apoptosis in animals given 0.5 mg ovalbumin. Furthermore, there was no increased deletion of CD4⁺, V β 8.2⁺ cells observed in animals fed 0.5 mg 20 times over a one-month period. *In situ* labelling of cells for degraded DNA demonstrated a large percentage (up to 10%) of cells undergoing programmed cell death in the dome area of Peyer's patches from mice fed 500 mg ovalbumin (Fig. 2f). This was not seen in control animals (Fig. 2d), and was only minimally present in animals fed low doses of ovalbumin (Fig. 2e).

Administration of superantigens may cause deletion, which is preceded by T-cell activation^{17,18}. After 500 mg ovalbumin was given once, the percentage of activated Peyer's patch T cells, as determined by being in the S/G2-M phase of the cell cycle, rose to 7% and returned to zero after five feedings, presumably as a result of deletion (Fig. 3a-c). An increase in the percentage of activated cells also occurred in animals given 0.5 and 5 mg ovalbumin, and was maximal after three feedings (Fig. 3a), although deletion (Fig. 1) was only observed in mice given 5 mg. In addition, approximately 18% of T cells in animals fed twice with 500 mg ovalbumin express the cell-surface activation marker CD44 with low or no CD45RB, demonstrating T-cell activation following oral antigen and that this activation precedes deletion.

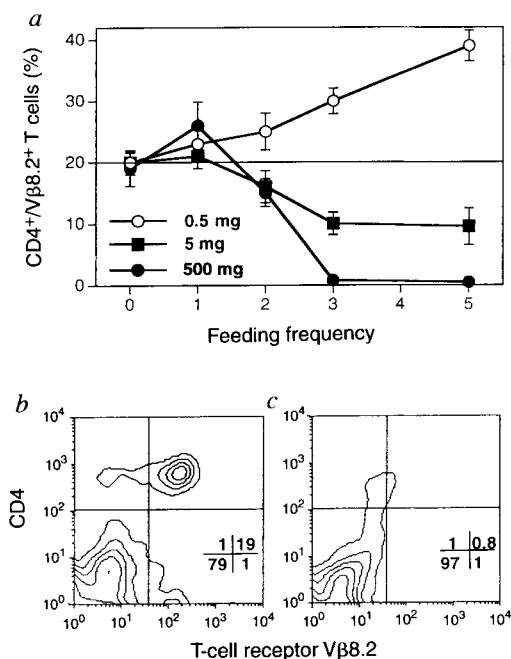
Previous studies have demonstrated that cells secreting TGF- β or Th2 cytokines (IL-4 and IL-10) are preferentially generated following oral administration of antigens^{19,20}. We found that with high-dose antigen feeding (500 mg) there was enhancement of secretion of both Th1 (IFN- γ) and Th2 (IL-4 and IL-10) cytokines in the spleen followed by complete loss of the production of these cytokines with continued feeding (Fig. 4). Secretion of IL-2 decreased without prior enhancement. A similar pattern was observed with 5 mg. In marked contrast, with low-dose antigen feeding (0.5 mg) there was progressive enhancement of the production of IL-4 and IL-10 with a minimal effect on the production of IL-2 and IFN- γ . These changes in Th1 and Th2 cytokine patterns are consistent with the observed increase in CD4⁺, V β 8.2⁺ T cells following low-dose (0.5 mg) feeding and the

decrease of these cells following high-dose (5 mg, 500 mg) feeding. However, all three feeding regimes enhanced the production of antigen-specific secretion of TGF- β . At the highest dose (500 mg), in addition to T cells, antigen-presenting cells also contributed to increased TGF- β production.

Antigen-specific hyporesponsiveness in mice fed 500 mg ovalbumin was observed even though there is a high frequency of ovalbumin-specific T cells in the transgenic animals. Splenic T-cell proliferative responses were suppressed by more than 90% in mice given ovalbumin compared with those that were not ($3,432 \pm 52$ versus $47,079 \pm 6,131$ Δ c.p.m.), and antibody responses were suppressed by 50 to 75% in mice fed 500 mg ovalbumin five times followed by subcutaneous immunization with 100 μ g ovalbumin and complete Freund's adjuvant. Anti-ovalbumin IgM titres were reduced from 512 ± 43 in non-fed to 128 ± 21 in fed mice; IgG1 was reduced from 32 ± 5 to 16 ± 7 ; and IgG2a was reduced from 256 ± 14 to 128 ± 36 . There were very few, if any, detectable anti-ovalbumin antibodies in naive transgenic mice. In addition to deletion, evidence of anergy was also observed in mice fed 500 mg ovalbumin. Specifically, the reduced splenic T-cell proliferative responses could be partially reversed (from $3,432 \pm 52$ to $24,227 \pm 1468$ Δ c.p.m.) by pre-culture of cells with recombinant IL-2.

Our finding that T-cell activation precedes deletion of T cells in the Peyer's patch suggests that, unlike thymic deletion, deletion following oral antigen administration is mediated by activation-induced apoptosis. Similar phenomena have been observed in peripheral deletion induced by superantigen^{17,18} and deletion of transgenic T cells specific for H-Y antigen²¹. It is possible that in some instances oral antigen only deletes T cells with high-affinity TCRs, leaving intact low-affinity T cells that respond poorly to antigen in the absence of IL-2. Whether these low-avidity TCR-bearing cells are anergized in oral tolerance is not clear. None the less, despite deletion of both Th1 and Th2 cells at very high doses, there was induction of cells producing TGF- β at all three doses of antigen. Thus cells producing TGF- β appear to be relatively resistant to peripheral deletion, perhaps because of the central role of TGF- β in promoting IgA synthesis

FIG. 1 Frequency of CD4⁺, V β 8.2⁺ T cells in Peyer's patch following antigen feeding. Ovalbumin-TCR transgenic mice were given 0.5 to 500 mg of ovalbumin every second day for a total of five feedings. Mice were killed either before the first feeding or 24 h after feeding. Peyer's patches (5 to 8 per mouse) were collected from the small intestine and a single-cell suspension was prepared²³. Cells were first centrifuged through a Ficoll-isopaque gradient and then stained for CD4 (with PE-conjugated YTS 191.1 mAb, Caltag, San Francisco) and V β 8.2 (with FITC-conjugated MR5-2 mAb, Pharmingen, San Diego). Fluorescence was analysed on a Becton-Dickinson FACSort using Lysis II software. Data collection was gated on live cells through propidium iodide exclusion, and data represent 10,000 events presented as probability (15%) contours. Each data point in a represents an average from 6 to 10 mice pooled from three independent experiments. a, Percentage of CD4⁺, V β 8.2⁺ T cells in the Peyer's patches. The actual numbers of CD4⁺, V β 8.2⁺ cells in Peyer's patches in a total of 10,000 cells counted were as follows: Peyer's patches of non-fed controls contained 1,921 CD4⁺, V β 8.2⁺ cells. Following high-dose feeding with 500 mg ovalbumin, it increased to 2,634 after the first feeding and fell to 50 by the fifth feeding. In mice given 5 mg there was a decrease in the cell number beginning after the second feeding and reaching 950 by the fifth feeding. In contrast, in mice given 0.5 mg the cell numbers steadily increased to 3,942 by the fifth feeding. The statistical significance (chi square analysis) of frequency differences of CD4⁺, V β 8.2⁺ T cells among the various groups was as follows: for those given 0.5 mg, $P < 0.001$ versus 5 mg, 500 mg and unfed after 3 and 5 feedings; for 5 mg and 500 mg, $P < 0.001$ versus unfed after 3 and 5 feedings. b, c, FACS contour plots of Peyer's patches from mice given PBS or OVA (500 mg) 5 times.



in the gut²², and these cells may represent a unique subpopulation of T helper cells (Th3?)²⁰.

In summary, we have found that the induction of peripheral tolerance following oral administration of antigen includes clonal deletion of antigen-reactive T cells, and that this affects both Th1 and Th2 subsets. These findings have important impli-

cations not only for understanding the basic mechanisms of immunoregulation and peripheral deletion, but for the therapeutic applications of oral tolerance as well. Furthermore they demonstrate that multiple mechanisms of tolerance have evolved to prevent the host from mounting a harmful immune response to ingested antigen, and that the mechanism triggered is depen-

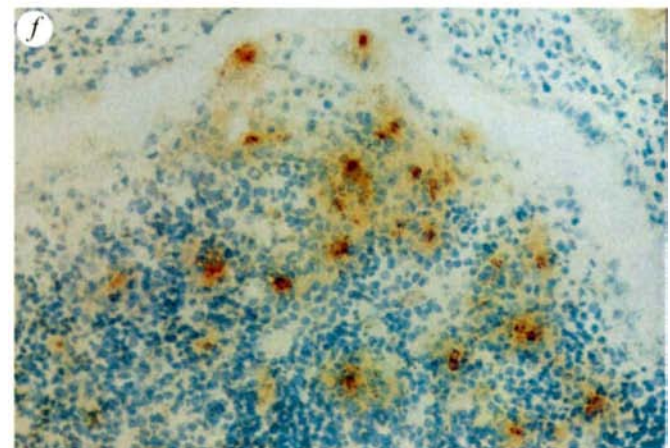
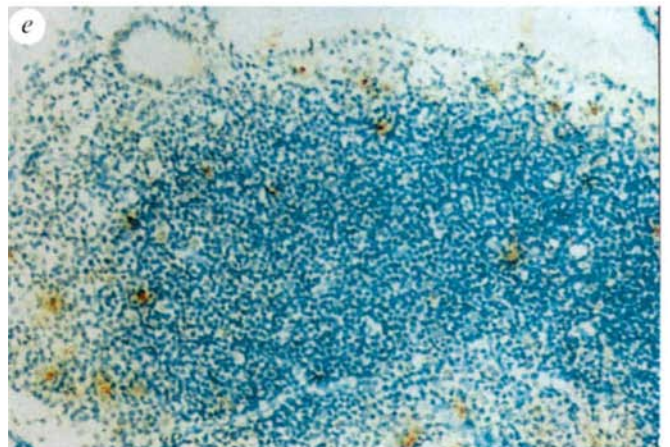
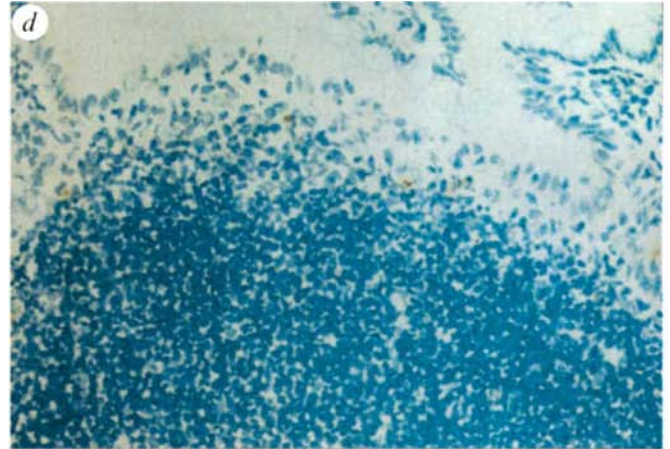
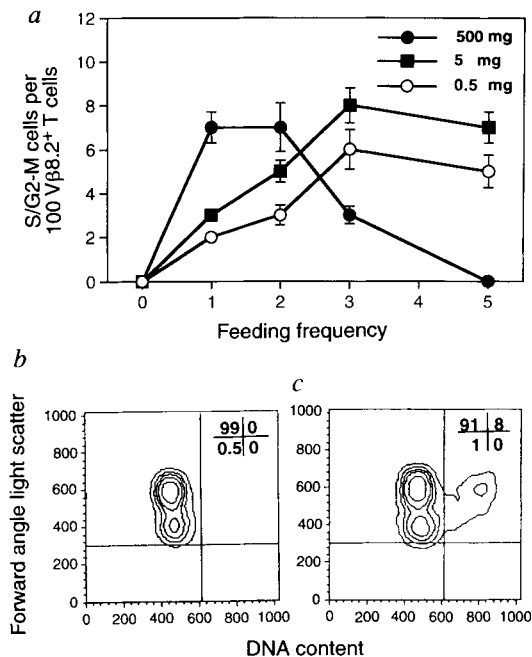


FIG. 2 Apoptosis of $V\beta 8.2^+$ T cells in Peyer's patch following antigen feeding. Peyer's patch cells as prepared in Fig. 1 were stained for $V\beta 8.2$ using FITC-conjugated MR5-2 mAb, and for degraded DNA using acridine orange (Sigma, St Louis). Data collection was gated on live $V\beta 8.2^+$ cells, and data represent 10,000 events presented as probability (15%) contours. Each data point in a represents an average from 6 to 10 mice pooled from three independent experiments. To demonstrate apoptosis, cells were stained with acridine orange using a modified version of a method described previously²⁴. Briefly, 10^6 Ficoll separated cells were first stained for $V\beta 8.2$ and then incubated in 100 μ l DMEM culture medium containing 10 μ g acridine orange at 25 °C for 15 min. Cells were then washed and examined directly without fixation. a, Number of apoptotic $V\beta 8.2^+$ T cells. The numbers of cells undergoing apoptosis in 10,000 $V\beta 8.2^+$ cells counted were as follows: Peyer's patches of non-fed control mice contained 0 to 50 cells undergoing apoptosis. Following high-dose feeding with 500 mg the numbers of apoptotic cells increased to a peak of 1,338 cells by the second feeding and decreased to 241 cells by the fifth feeding. In mice given 5 mg the number of apoptotic cells gradually increased and reached to a peak of 803 cells by the fifth feeding. In mice given 0.5 mg very small numbers of cells (53 to 114) undergoing apoptosis were observed. The statistical significance (chi square analysis) of the frequency differences of apoptotic $V\beta 8.2^+$ T cells was as follows: for 500 mg, $P < 0.001$ after 2 feedings versus 5 mg, 0.5 mg and unfed; for 5 mg, $P < 0.01$ versus 0.5 mg and unfed on days 3 and 5; b, c, FACS contour plots of Peyer's patches from mice given PBS or ovalbumin (500 mg) twice. d–f, *In situ* labelling of apoptotic cells in the Peyer's patch. Peyer's patches were frozen in OCT compound (Miles, Naperville, IL), and cryostat sectioned (6 μ m). For direct labelling of degraded DNA, the Oncor ApopTag system was used according to manufacturer's instructions (Oncor, Gaithersburg, MD). Residues of dioxigenin-nucleotide are added to the DNA by terminal deoxynucleotidyl transferase and newly synthesized nucleotide polymers were then detected by peroxidase-labelled anti-digoxigenin antibody and the chromogen diaminobenzidine. The dome area of Peyer's patches from ovalbumin TCR transgenic mice 24 h after the second feeding with d, 1 mg HEL as control (magnification, $\times 78$), e, 1 mg ovalbumin ($\times 63$) and f, 500 mg ovalbumin ($\times 94$) is shown.

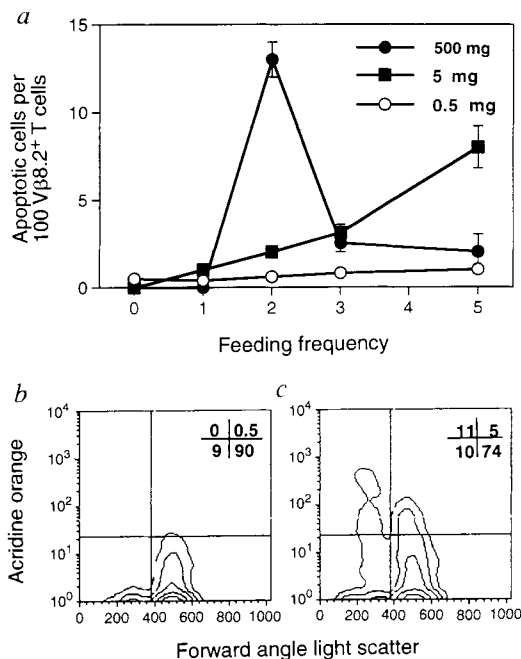
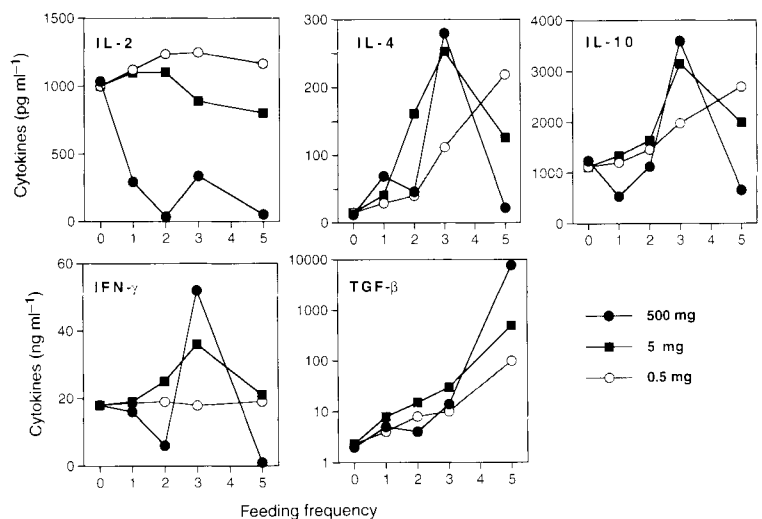


FIG. 3 T-cell activation in Peyer's patch following antigen feeding. Peyer's patch cells as prepared in Fig. 1 were stained for V β 8.2 with FITC-conjugated MR5-2 mAb and for total DNA with propidium iodide²⁵. Data collection was gated on live V β 8.2⁺ cells, and data represent 10,000 events presented as probability (15%) contours. Each data point in a represents an average from 6 to 10 mice pooled from three independent experiments. a, Number of V β 8.2⁺ T cells at S/G2-M phases of cell cycle. The numbers of cells in S/G2-M phase of cell cycle in 10,000 V β 8.2⁺ cells counted were as follows: Peyer's patches of non-fed control mice contained 0 to 45 V β 8.2⁺ cells in S/G2-M phase. Following high-dose feeding with 500 mg the S/G2-M cells increased to 714 and 729 by the first and second feeding respectively, and decreased to an undetectable level by the fifth feeding. In mice given 5 mg the number of cells in S/G2-M phase steadily increased to 826 by the third feeding and was 735 by the fifth feeding. In mice given 0.5 mg the number of cells in S/G2-M phase increased to 641 and was 528 cells by the fifth feeding. The statistical significance (chi square analysis) of the frequency differences of S/G2-M, V β 8.2⁺ T cells was as follows: for 500 mg, $P < 0.001$ after one to three feedings versus unfed; for 0.5 and 5 mg, $P < 0.01$ versus unfed at all time points. b, c, FACS contour plots of Peyer's patches from animal fed once with PBS or ovalbumin (500 mg).

FIG. 4 Oral administration of antigen activates and tolerizes both Th1 and Th2 cells in a dose-dependent fashion. Ovalbumin-TCR transgenic mice were fed and killed as described in Fig. 1. Splenocytes, 4×10^5 cells per well, were cultured in 0.2 ml serum-free medium containing various concentrations of ovalbumin. Peyer's patches were not used because they contained an inadequate number of cells for multiple cytokine assays. Culture supernatants were collected after 40 h (for IL-2, IL-4, IL-10 and IFN- γ) or 72 h (for TGF- β). Cytokine concentration was determined by ELISA. Quantitative ELISA for IL-2, IL-4, IL-10 and IFN- γ was performed using paired mAbs specific for corresponding cytokines according to manufacturer's recommendations (Pharmingen, San Diego). Active TGF- β 1 (without acid treatment) was determined by a sandwich ELISA as described¹⁴. Data presented represent mean of cultures with 1 mg ml^{-1} ovalbumin minus mean of cultures with 1 mg ml^{-1} HEL. Each data point represents an average from 4–6 mice; the standard deviation is within 15% of the mean. The experiment was repeated three times with similar results. The statistical significance (as determined by Student's t test) for ovalbumin fed versus unfed was as follows: $P < 0.0001$ for IL-2 after feeding 500 mg ovalbumin 2–5 times; $P < 0.001$ for IL-4 and IL-10 after 5 mg or 500 mg ovalbumin 3 times; $P < 0.001$ for IL-4 and IL-10 after 0.5 mg ovalbumin 5 times. $P < 0.001$ for IFN- γ after 5 mg or 500 mg ovalbumin 3 times; $P < 0.0001$ for TGF- β after feeding with ovalbumin 3–5 times at all doses.



dent upon the amount of antigen encountered by the mucosal immune system. □

Received 25 October 1994; accepted 16 May 1995.

- Wells, H. G. *J. infect. Dis.* **8**, 147–171 (1991).
- Chase, M. *Proc. Soc. exp. Biol. Med.* **61**, 257–259 (1946).
- Mowat, A. M. *Immun. Today* **8**, 93–98 (1987).
- Thompson, H. S. G. & Staines, N. A. *Clin. exp. Immun.* **64**, 581–586 (1986).
- Nagler-Anderson, C., Bober, L. A., Robinson, M. E., Siskind, G. W. & Thorbecke, F. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7443–7446 (1986).
- Bitar, D. M. & Whitacre, C. C. *Cell. Immun.* **112**, 364–370 (1988).
- Higgins, P. & Weiner, H. L. *J. Immun.* **140**, 440–445 (1988).
- Zhang, J. A., Davidson, L., Eisenbarth, G. & Weiner, H. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10252–10256 (1991).
- Nussenblatt, R. B. et al. *J. Immun.* **144**, 1689–1695 (1990).
- Weiner, H. L. et al. *Rev. Immun.* **12**, 809–837 (1994).
- Weiner, H. L. et al. *Science* **259**, 1321–1324 (1993).
- Trentham, D. E. et al. *Science* **261**, 1727–1730 (1993).
- Whitacre, C. G., Gienapp, I. E., Orosz, C. G. & Bitar, D. J. *Immun.* **147**, 2155–2163 (1991).

- Friedman, A. & Weiner, H. L. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6688–6692 (1994).
- Murphy, K., Heimberger, A. & Loh, D. *Science* **250**, 1720–1723 (1990).
- Huang, L., Sye, K. & Crispe, I. N. *Int. Immun.* **6**, 533–540 (1994).
- Webb, S., Morris, C. & Sprent, J. *Cell* **63**, 1249–1256 (1990).
- Kawabe, Y. & Ochi, A. *Nature* **349**, 245–247 (1991).
- Xu-Amano, J., Aicher, W. K., Taguchi, T., Kiyono, H. & McGhee, J. R. *Int. Immun.* **4**, 433–445 (1992).
- Chen, Y., Kuchroo, V. K., Inobe, J.-I., Hafler, D. A. & Weiner, H. L. *Science* **265**, 1237–1240 (1994).
- Rocha, B. & Boehmer, H. V. *Science* **251**, 1225–1228 (1992).
- Coffman, R., Leberman, D. & Shrader, B. J. *exp. Med.* **144**, 3411–3416 (1989).
- Santos, L. M. B., Al-Sabbagh, A., Londono, A. & Weiner, H. L. *Cell. Immun.* **157**, 439–447 (1994).
- Hardin, J., Sherr, D., DeMaria, M. & Lopez, P. J. *immunol. Meth.* **154**, 99–107 (1992).
- Noguchi, P. D. in *Current Protocols in Immunology* (eds Coligan, J. E., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M. & Strober, W.) 5.7.1–5.7.3 (Wiley, Secaucus, New York, 1994).

ACKNOWLEDGEMENTS. We thank D. Y. Loh for providing the ovalbumin-TCR transgenic mice. This work was supported by grants from the N.I.H. Y.C. is a recipient of a MRC fellowship from Canada.

CD95 (Fas)-dependent elimination of self-reactive B cells upon interaction with CD4⁺ T cells

Jeffrey C. Rathmell*, Michael P. Cooke†‡, William Y. Ho§, Jeff Grein†, Sarah E. Townsend†, Mark M. Davis†§ & Christopher C. Goodnow†§¶

* Program in Immunology, † Howard Hughes Medical Institute, and § Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA

‡ Present address: Systemix Inc., 3155 Porter Avenue, Palo Alto, California 94306, USA

¶ To whom correspondence should be addressed

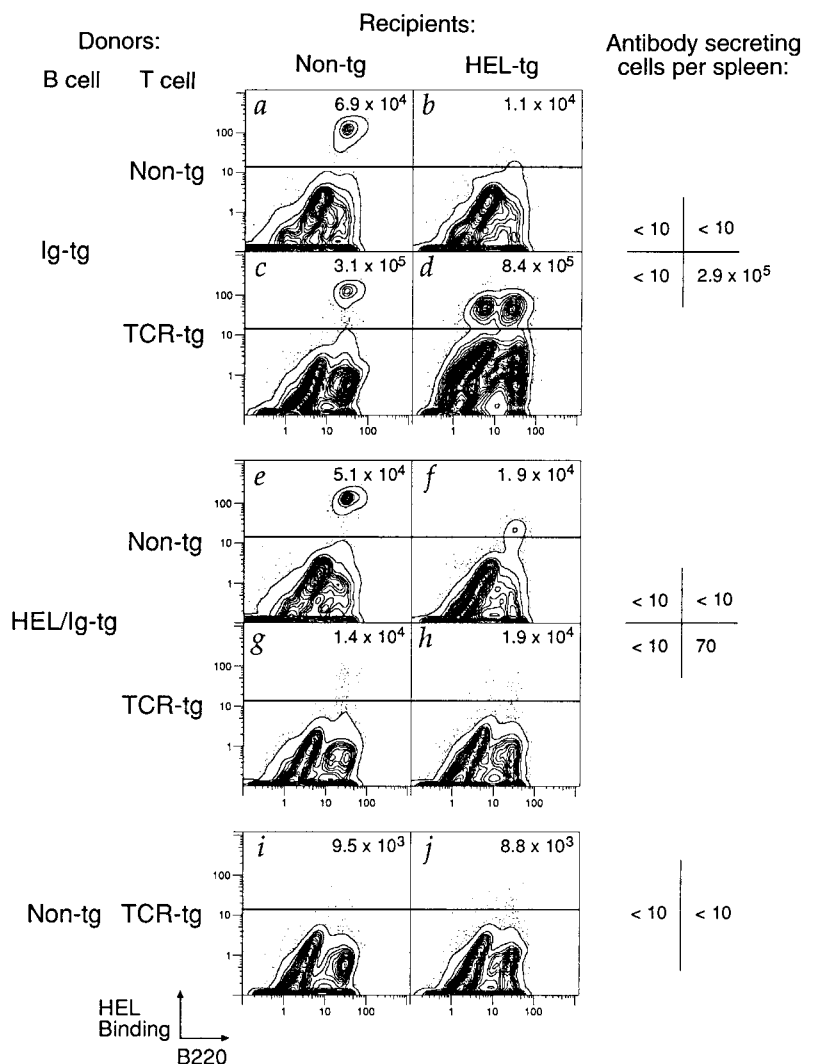
THE recessive mouse mutations *lpr* and *gld* create deficiencies in an interacting pair of cell surface molecules, CD95 (Fas/APO-1) and Fas-ligand (FasL), respectively¹⁻³, resulting in autoantibody production resembling human systemic lupus erythematosus⁴. The

mechanisms of self-tolerance affected by deficiency in either molecule are not established, but CD95 deficiency both in B cells and in CD4⁺ T cells recognizing major histocompatibility complex (MHC) class II molecules is required for autoimmunity in *lpr* mice⁵⁻⁸. Here we track the outcome of *in vivo* interactions between B cells and CD4⁺ T cells that recognize a transgene-encoded autoantigen, hen egg lysozyme (HEL), using cells from mice transgenic for immunoglobulin and T-cell receptor (TCR) genes. B cells that had not previously encountered HEL autoantigen (naïve cells) were triggered into proliferation and antibody production upon interaction with antigen and HEL-specific CD4⁺ T cells. By contrast, B cells that had been chronically exposed to HEL during their development and carried desensitized surface immunoglobulin (sIg) antigen receptors⁹ (anergic cells) did not produce antibody but instead were eliminated in the presence of HEL-specific CD4⁺ T cells. CD95-deficient anergic B cells, however, were not eliminated by CD4⁺ T cells and were triggered to proliferate. These findings identify a novel regulatory step for eliminating autoreactive B cells that seems unique in its dependence on CD95.

To study interactions between autoreactive B and T cells *in vivo*, B and T cells from mice transgenic for HEL-specific immunoglobulin^{9,10} and TCR¹¹ respectively, were mixed, transferred into irradiated histocompatible recipients that either did or did not express HEL as an autoantigen (HEL-transgenic and

FIG. 1 Outcome of *in vivo* interaction between naïve or anergic B cells and T cells specific for HEL autoantigen. *a-j*, Flow cytometric analysis of spleen cells in adoptive transfer recipients 5 days after transferring cell suspensions containing either naïve HEL-specific B cells from mice transgenic for immunoglobulin (Ig-tg; *a-d*), anergic HEL-specific B cells from HEL/Ig transgenic mice (HEL/Ig-tg; *e-h*), or B cells that were not HEL-specific from non-transgenic mice (Non-tg; *i, j*). The transferred cell mixtures also contained an approximately equal number of HEL-specific T cells from TCR-transgenic mice (TCR-tg; *c, d, g, h, i, j*) or T cells that were not HEL-specific from non-transgenic mice (Non-tg; *a, b, e, f*). The recipients lacked (Non-tg; *a, c, e, g, i*) or expressed HEL autoantigen (HEL-tg; *b, d, f, h, j*) as a secreted serum protein. Transferred B cells were enumerated in recipient spleen cell suspensions by three-colour immunofluorescent staining for HEL-binding receptors (HEL), a pan-B cell marker (B220), and a plasma cell-specific marker (syndecan; not shown). The population of B220^{low} HEL-binding cells in *d* stain brightly for syndecan (not shown) and represent plasmablasts. Equivalent data were obtained for duplicate animals in each experimental group, and the mean number of HEL-binding B220⁺ cells per spleen in each group is shown in the upper right of each panel. The apparent increase in HEL-specific B cells in *c* relative to *a* was not observed in other experiments and was not accompanied by blastogenesis. The column on the right shows the mean number of anti-HEL IgM⁹ antibody-secreting cells per spleen in the same groups. Comparable results were obtained in four separate experiments. In one experiment, blood and mesenteric, cervical and inguinal lymph nodes were also analysed and gave comparable results to *e* and *g*.

METHODS. All donors and recipients were (C57BL/6J × B10.BR/SgSnJ)_{F1}-hybrid mice, sex-matched, and between 6 and 20 weeks of age. The donor cell inoculum consisted of 10⁷ splenocytes from MD4 immunoglobulin-transgenic, ML5/MD4 HEL/immunoglobulin double-transgenic¹⁰, or non-transgenic mice mixed with 10⁷ splenocytes from either non-transgenic or 3A9 TCR transgenic¹¹ animals. Donor cells were injected into the lateral tail vein of sublethally X-irradiated (750 rad) non-transgenic or ML5 HEL transgenic recipients. Five days later, spleen cell suspensions from the recipient mice were analysed by immunofluorescent staining and flow cytometry as described previously¹⁷, using streptavidin-CyChrome (Pharmingen). Antibody-secreting cells were enumerated by spot enzyme-linked immunosorbent assay as previously described¹⁷. Absolute numbers of HEL-



binding B cells were determined by multiplying the frequency of HEL-binding B220⁺ cells by the number of spleen cells determined by hemacytometer.

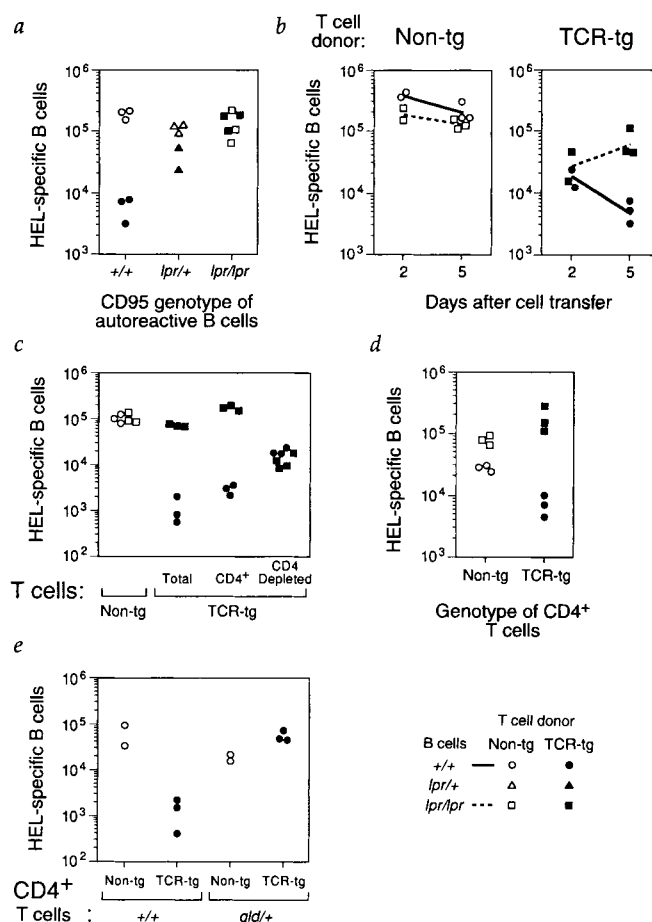
non-transgenic), and tracked flow cytometrically (Fig. 1). Naive HEL-specific B cells from mice transgenic for immunoglobulin became activated upon exposure to HEL autoantigen (ref. 9 and results not shown) and they proliferated and differentiated into antibody-secreting cells when TCR-transgenic cells were also present (Fig. 1d and corresponding quadrant in right column). By contrast, anergic B cells from HEL/immunoglobulin double-transgenic mice, which fail to become activated by HEL autoantigen⁹, did not proliferate or differentiate into antibody-secreting cells in the presence of HEL antigen and HEL-specific T cells (Fig. 1h and corresponding quadrant in right column). Moreover, the anergic B cells selectively disappeared from spleen, lymph nodes, and blood in non-transgenic recipients when specific TCR-transgenic T cells were added (Fig. 1g and results not shown). This result suggested that the HEL-peptides constitutively presented by anergic self-reactive B cells^{11,12} were recognized by the MHC class-II-restricted TCR on the T cells, but that in this context the T cells brought about B-cell death rather than activation.

To explore the possibility that T-cell-dependent elimination

of anergic B cells was mediated through CD95, the *in vivo* transfers were repeated with CD95-deficient (anergic) B cells from young HEL/immunoglobulin double-transgenic *lpr* mice (Fig. 2a). Anergic B cells carrying wild-type CD95 genes (+/+) decreased 20-fold in number in the presence of HEL-specific TCR-transgenic cells (filled circles), whereas anergic B cells that were homozygous for the *lpr* mutation were not fewer in number 5 days after transfer with TCR-transgenic cells (*lpr/lpr*; filled squares) and *lpr* heterozygous B cell numbers decreased only fourfold (*lpr/+*, filled triangles). Two separate mechanisms mediated the TCR-transgenic T-cell effect, because when anergic HEL-binding B cells were enumerated 2 days after transfer (Fig. 2b) +/+ and *lpr/lpr* cells both decreased approximately 10-fold in the presence of TCR-transgenic cells. By day 5, however, the *lpr* B cells numbers had increased whereas the +/+ self-reactive B cells continued to disappear. Separation of the TCR-transgenic cells into CD4⁺ and CD4-depleted fractions (Fig. 2c) revealed that the CD4⁺ cells triggered elimination of +/+ anergic B cells efficiently but caused *lpr/lpr* anergic B cells to double in number by day 5. By contrast, transferring anergic B

FIG. 2 Anergic self-reactive B cells resist elimination by HEL autoantigen-specific CD4⁺ T cells when the B cells are deficient in CD95 or the T cells are deficient in FasL. a, Numbers of HEL-binding B cells in individual recipient spleens 5 days after adoptive transfer of anergic HEL-specific B cells that carried wild-type CD95 genes (+/+; circles), or were heterozygous (*lpr/+*; triangles) or homozygous for the *lpr* mutation (*lpr/lpr*; squares). The B cells were transferred into non-transgenic recipients, as in Fig. 1, together with TCR transgenic (filled symbols) or non-transgenic T cells (open symbols). b, Numbers of HEL-binding B cells in individual recipients as in a, except that recipients were analysed 2 and 5 days after cell transfer. Geometric means are indicated for +/+ (solid lines) and *lpr/lpr* (dotted lines) HEL/immunoglobulin transgenic B cells. c, Numbers of HEL-binding B cells in individual transfer recipients 5 days after transfer as in a, except that additional groups received TCR-transgenic cells fractionated into CD4⁺ or CD4-depleted cells. The number of V β 8.2⁺ CD4⁺ or CD4⁻ cells given was equal to the numbers of cells in these subsets contained in the unfractionated mixture. Equivalent results were obtained in 3 separate experiments with fractionated T cells. d, Numbers of HEL-binding B cells 5 days after cell transfer as in a, except that the anergic B cells were mixtures of Ly5^a-positive +/+ and Ly5^a-negative *lpr/lpr* cells obtained from the spleen of a mixed-bone-marrow chimera in which the two types of B cells had developed side by side. The transfer inoculum contained a 2:1 ratio of *lpr/lpr* Ly5^a- B cells to Ly5^a+ B cells, together with CD4⁺ T cells. e, Numbers of HEL-binding B cells 5 days after cell transfer, as in a, except with CD4⁺ T cells from *gld/+* or +/+ TCR-transgenic or non-transgenic mice. Heterozygous *gld/+* mice were used as a source of FasL-deficient T cells because: (i) the T cells had not been subject to potential secondary effects of developing in the autoimmune environment typical of *gld* homozygotes; (ii) epistatic interaction between *gld* and *lpr*^{cg} in *gld/+lpr^{cg}/+* mice indicates that functional FasL is diminished in *gld/+* animals^{4,3}; (iii) CD95 itself was limiting during anergic B-cell elimination based on the partial resistance of *lpr/+* B cells (Fig. 2a).

METHODS. Donor and recipients were as in Fig. 1, except that the *lpr* mutation from second- and third-generation backcrosses of B10.BR/SgSnJ onto the C57BL/6J-*lpr/lpr* background was introduced into the transgenic B-cell donors. Mice were screened by PCR for H-2^{k/b} MHC haplotype heterozygosity and the *lpr* mutation²⁸. They were confirmed as *lpr* homozygotes or heterozygotes by staining thymocytes for CD95 expression using the Jo2 antibody (provided by S. Nagata). *gld/+* animals were from crosses between H-2^{k/b} TCR-transgenic +/+ animals and C57BL/6J-*gld/gld* mice. The number of cells transferred from B- and T-cell donors was 1.35×10^7 cells from each in a, and 1.5×10^7 from each in b. In c, 10^7 splenocytes from B-cell donors were transferred with 6.5×10^6 cells from T-cell donors or with the number of V β 8.2⁺ CD4⁺ or V β 8.2⁺ CD4⁻ cells present in 6.5×10^6 unfractionated cells. In d, mixed-bone-marrow chimera were generated by reconstituting lethally irradiated HEL-transgenic (B10.BR $\times$ C57BL/6J)F₁ mice with 2.5×10^6 +/+ Ly5^a+ and 2.5×10^6 *lpr/lpr* Ly5^a- anti-HEL immunoglobulin-transgenic bone marrow cells, and splenocytes containing both types of B cell were collected 28 days after reconstitution. In e, 10^7 splenocytes from B-cell donors were transferred with 5×10^5 CD4⁺



T-cell enriched splenocytes. CD4⁺ cells were enriched using anti-CD4-microbeads and one or two cycles of binding to a MiniMacs column (Miltenyi Biotec, Gladbach, Germany) as described by the manufacturer, and were 66.7% and 83.4% CD4⁺ in c and d, respectively, and 98.7%, 97.0%, 88.8% and 93.3% CD4⁺ for the +/+ non-transgenic, +/+ TCR-transgenic, *gld/+* non-transgenic, and *gld/+* TCR-transgenic T cells, respectively. CD4-depleted cells were the flow-through following five cycles over the MiniMacs column and contained <0.01% CD4⁺ cells. HEL-binding B cells were enumerated in recipient spleens as in Fig. 1 except that B220⁺ IgM^a HEL-binding cells were counted. In a-e the lower limits for detecting HEL-binding cells were 3.6×10^3 , 8.0×10^3 , 1.6×10^3 , 2.1×10^3 and 6.4×10^2 B cells per spleen, respectively.

cells with CD4-depleted TCR-transgenic cells resulted in the CD95-independent decrease in $+/+$ and lpr/lpr B-cell numbers, demonstrating the necessity for CD4 $^{+}$ T cells in the CD95-dependent elimination of anergic B cells.

The possibility that lpr/lpr anergic B cells resisted T-cell-dependent elimination because of an indirect effect of CD95 deficiency was excluded by two observations. First, mixtures of $Ly5^{a+}+/+$ and $Ly5^{a-}lpr/lpr$ anergic B cells that had developed side by side in chimaeric animals retained their respective susceptibility or resistance to killing upon transfer with CD4 $^{+}$ TCR-transgenic T cells (Fig. 2d). Second, $gld/+$ TCR-transgenic CD4 $^{+}$ T cells, which have limited functional FasL (ref. 13 and see Fig. 2 legend) but were not subject to potential secondary effects of developing in the autoimmune environment typical of gld/gld animals, were unable to eliminate $+/+$ anergic B cells (Fig. 2e). Taken together, susceptibility or resistance to CD4 $^{+}$ T-cell-dependent killing was therefore determined by the ability of the anergic B cell to express CD95 and by the T cell's capacity to express FasL.

The selective increase in lpr/lpr B cell numbers after transfer with HEL-specific CD4 $^{+}$ T cells (Fig. 2b-d) suggested that CD95-deficient anergic B cells proliferated upon interaction with the T cells instead of being eliminated. In confirmation of this notion, anergic lpr B cells transferred with TCR-transgenic cells

regained normal IgM expression and had reduced IgD, which is consistent with loss of tolerance and activation (Fig. 3a, b); became large blast cells (Fig. 3c); increased expression of CD44 (not shown); and incorporated bromodeoxyuridine (BrdU; Fig. 3d). Anergic lpr B cells did not proliferate when CD4-depleted T cells were used (results not shown). Interestingly, no differentiation into antibody-secreting cells accompanied the CD4 $^{+}$ T-cell-dependent proliferation of anergic lpr B cells (results not shown), in contrast to the vigorous antibody production by naive B cells (Fig. 1).

This novel mechanism for eliminating self-reactive B cells appears unique among B-cell censoring mechanisms identified so far in its dependence on CD95, because autoreactive B-cell elimination in the bone marrow^{14,15} or by follicular exclusion¹⁶ are not affected by B-cell deficiency in CD95 (ref. 17 and J.C.R., C.C.G. and J. Cyster, unpublished data). The desensitization of surface immunoglobulin signalling that underlies B-cell anergy⁹ is also unaffected by B-cell CD95 deficiency¹⁷, although the functional tolerance that results from surface-immunoglobulin-receptor desensitization is shown here to be reinforced by CD95-dependent elimination of anergic cells. On the basis of the data above and detailed studies of CD95 and FasL expression and function in other systems^{1,3,18,19}, we suggest that autoantigens presented by anergic B cells trigger FasL expression on specific T cells and induce B-cell apoptosis by engaging CD95 on the B cell. Anergic B cells may be selectively vulnerable to CD95-dependent cytotoxicity because their prior chronic exposure to HEL autoantigen desensitizes surface immunoglobulin signalling⁹, rendering them either unable to display co-stimulatory molecules that trigger T cell cytokines such as IL4 which suppress B cell apoptosis and favour mitosis^{9,11}, and/or unable to transmit surface immunoglobulin signals that may internally suppress CD95-induced apoptosis^{20,21}. Previous studies have shown that CD95 deficiency in either B cells or T cells alone is insufficient for the high level of autoantibody production that characterizes lpr mice^{5,8}. Although resistance to this censoring mechanism may be the primary B-cell defect in C57BL/6- lpr mice^{5,6}, the observation that anergic lpr B cells did not secrete autoantibody despite being triggered into proliferation by $+/+$ CD4 $^{+}$ T cells indicates that additional defects must be present. These could include the inappropriate expression of costimulatory molecules on anergic B cells^{9,11,21} and the presence of activated autoreactive T cells that have escaped regulatory mechanisms that depend on CD95 expression by T cells^{7,19,21-28}. □

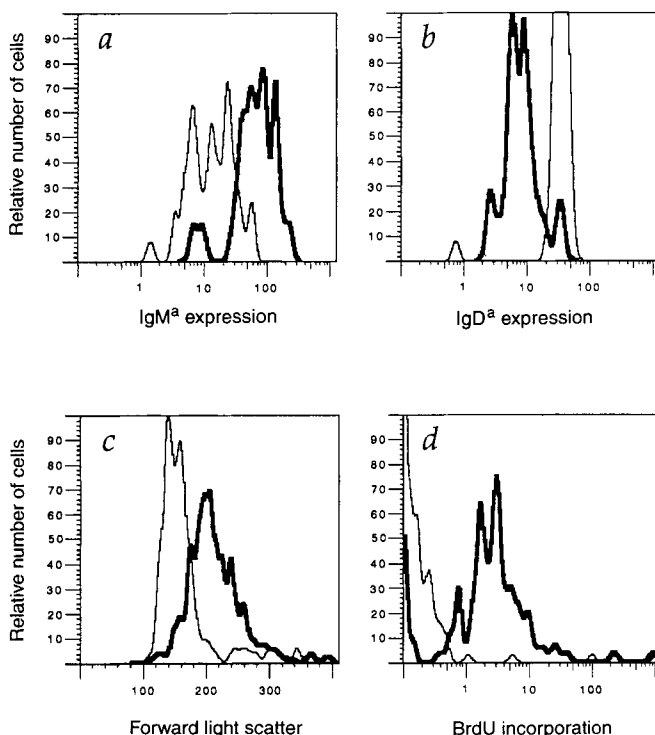


FIG. 3 Anergic lpr B cells recover IgM receptor expression, enlarge, and enter cell cycle upon interaction with HEL-autoantigen-specific CD4 $^{+}$ T cells. a-d, Anergic B cells from HEL/immunoglobulin-transgenic lpr mice were transferred with either TCR-transgenic T cells (thick histograms) or non-transgenic T cells (thin histograms), as in Fig. 2. Five days after transfer, HEL-binding cells were analysed by flow cytometry for cell-surface IgM a (a), cell surface IgD a (b), cell size as indicated by forward scattering of laser light (c) or for incorporation of BrdU into nuclear DNA (d). Incorporation of BrdU was measured in recipient animals fed with BrdU in the drinking water from days 2 to 5.

METHODS. B and T cells were transferred as in Fig. 2c and analysed for HEL-binding, IgM a , IgD a and cell size by flow cytometry as described previously¹⁷. BrdU (0.25 mg ml $^{-1}$) was added to the drinking water on day 2 after cell transfer and replaced daily. Two-colour immunofluorescent staining for cell-surface HEL-binding receptors and nuclear BrdU was performed as described²⁹.

Received 5 December 1994; accepted 16 May 1995.

1. Watanabe-Fukunaga, R., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Nagata, S. *Nature* **356**, 314-317 (1992).
2. Takahashi, T. et al. *Cell* **76**, 969-976 (1994).
3. Lynch, D. H. et al. *Immunity* **1**, 131-136 (1994).
4. Cohen, P. L. & Eisenberg, R. A. *Rev. Immun.* **9**, 243-269 (1991).
5. Nemazee, D., Guet, C., Buerki, K. & Marshak-Rothstein, A. *J. Immun.* **147**, 2536-2539 (1991).
6. Sobel, E. S. et al. *J. exp. Med.* **173**, 1441-1449 (1991).
7. Sobel, E. S., Cohen, P. L. & Eisenberg, R. A. *J. Immun.* **150**, 4160-4167 (1993).
8. Jevnikar, A. M., Grusby, M. J. & Glimcher, L. H. *J. exp. Med.* **179**, 1137-1143 (1994).
9. Cooke, M. P. et al. *J. exp. Med.* **179**, 425-438 (1994).
10. Goodnow, C. C. et al. *Nature* **334**, 676-682 (1988).
11. Ho, W. Y., Cooke, M. P., Goodnow, C. C. & Davis, M. M. *J. exp. Med.* **179**, 1539-1549 (1994).
12. Kanost, D. & McCluskey, J. *Eur. J. Immun.* **24**, 1186-1193 (1994).
13. Matsuzawa, A. et al. *J. exp. Med.* **171**, 519-531 (1990).
14. Nemazee, D. A. & Bürki, K. *Nature* **337**, 562-566 (1989).
15. Hartley, S. B. et al. *Nature* **353**, 765-768 (1991).
16. Cyster, J. G., Hartley, S. B. & Goodnow, C. C. *Nature* **371**, 389-395 (1994).
17. Rathmell, J. C. & Goodnow, C. C. *J. Immun.* **153**, 2831-2842 (1994).
18. Trauth, B. C. et al. *Science* **245**, 301-305 (1989).
19. Nagata, S. & Golstein, P. *Science* **267**, 1449-1456 (1995).
20. Rothstein, T. L. et al. *Nature* **374**, 163-165 (1995).
21. Goodnow, C. C. et al. *Adv. Immun.* **59**, 279-368 (1995).
22. Weston, K. M., Ju, S., Lu, C. Y. & Sy, M. J. *Immun.* **141**, 1941-1948 (1988).
23. Zhou, T. et al. *J. Immun.* **147**, 466-474 (1991).
24. Zhou, T., Bluthmann, H., Zhang, J., Edwards, C. K. & Mountz, J. D. *J. exp. Med.* **176**, 1063-1072 (1992).
25. Bossu, P. et al. *J. Immun.* **151**, 7233-7239 (1993).
26. Russell, J. H., Rush, B., Weaver, C. & Wang, R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4409-4413 (1993).

27. Gillette-Ferguson, I. & Sidman, C. L. *Eur. J. Immun.* **4**, 1181–1185 (1994).
 28. Singer, G. G. & Abbas, A. K. *Immunity* **1**, 365–371 (1994).
 29. Fulcher, D. A. & Basten, A. J. *exp. Med.* **179**, 125–134 (1994).

ACKNOWLEDGEMENTS. We thank M. Seldin and L. Berg for providing, respectively, sequences for PCR screening *Ipr* and MHC haplotypes, the Stanford Shared FACS Facility for flow cytometric support, and the Division of Laboratory Animal Medicine for animal husbandry. J.C.R. was supported by an N.S.F. predoctoral fellowship and W.Y.H. by an NIH training grant to the M.S.T.P. This research was supported by a grant from the Searle Scholars/Chicago Community Trust to C.C.G. and by the HHMI, C.C.G. and M.M.D. are investigators of the Howard Hughes Medical Institute.

A giant nucleopore protein that binds Ran/TC4

N. Yokoyama^{*†}, N. Hayashi^{*}, T. Seki^{*},
 N. Panté^{||}, T. Ohba^{*}, K. Nishii[‡],
 K. Kuma[§], T. Hayashida^{*}, T. Miyata[§],
 U. Aebi[¶], M. Fukui[†] & T. Nishimoto^{*#}

^{*} Department of Molecular Biology, [†] Department of Neurosurgery,
[‡] Department of Anatomy, Graduate School of Medical Science,
 Kyushu University, Maidashi 3-1-1, Higashi-ku,
 Fukuoka 812-82, Japan

[§] Department of Biophysics, Faculty of Science, Kyoto University,
 Sakyo-ku, Kyoto 606, Japan

^{||} M. E. Müller Institute for Microscopy, Biozentrum,
 University of Basel, CH-4056 Basel, Switzerland

[¶] Department of Cell Biology and Anatomy,
 The Johns Hopkins University School of Medicine, Baltimore,
 Maryland 21205, USA

RAN/TC4 is a small nuclear G protein¹ that forms a complex with the chromatin-bound guanine nucleotide release factor RCC1 (ref. 2). Loss of RCC1 causes defects in cell cycle progression^{3,4}, RNA export^{5–7} and nuclear protein import⁸. Some of these can be suppressed by overexpression of Ran/TC4 (ref. 1), suggesting that Ran/TC4 functions downstream of RCC1. We have searched for proteins that bind Ran/TC4 by using a two-hybrid screen, and here we report the identification of RanBP2, a novel protein of 3,224 residues. This giant protein comprises an amino-terminal 700-residue leucine-rich region, four RanBP1-homologous (refs 9, 10) domains, eight zinc-finger motifs similar to those of NUP153 (refs 11, 12), and a carboxy terminus with high homology to cyclophilin¹³. The molecule contains the XFXFG pentapeptide motif characteristic of nuclear pore complex (NPC) proteins¹⁴, and immunolocalization suggests that RanBP2 is a constituent of the NPC. The fact that NLS-mediated nuclear import can be inhibited by an antibody directed against RanBP2 supports a functional role in protein import through the NPC.

We set up a two-hybrid screen in which human Ran/TC4 fused to the DNA-binding domain on GAL4 was used as a 'bait' to isolate complementary DNA clones from a library¹⁵ of human cDNA tagged at the GAL4-activation domain, by selection for His⁺ prototrophs followed by β -galactosidase assay. A total of 81 cDNA clones were isolated and classified into 10 independent clones based on their nucleotide sequences (Table 1). Except for H2 clone, which encodes human RanBP1 (ref. 10), all clones were isolated at least twice, indicating that our screen was saturated. Seven out of ten clones covered 5.1 kilobases (kb) of a continuous open reading frame (Fig. 1A) that derived from the 3' half of a messenger RNA of about 10 kb as judged by northern blotting (Fig. 2a). Using the most 5' clone (G5) of the obtained cDNAs as a probe of another human cDNA library, we isolated a cDNA of 10 kb. The extreme 5' end of the mRNA was isolated by polymerase-chain-reaction amplification of HeLa-cell mRNA, using a primer based on the nucleotide sequence located

TABLE 1 Interaction of isolated proteins with Ran/TC4: quantitative assay of β -galactosidase activity from the *lacZ* reporter construct

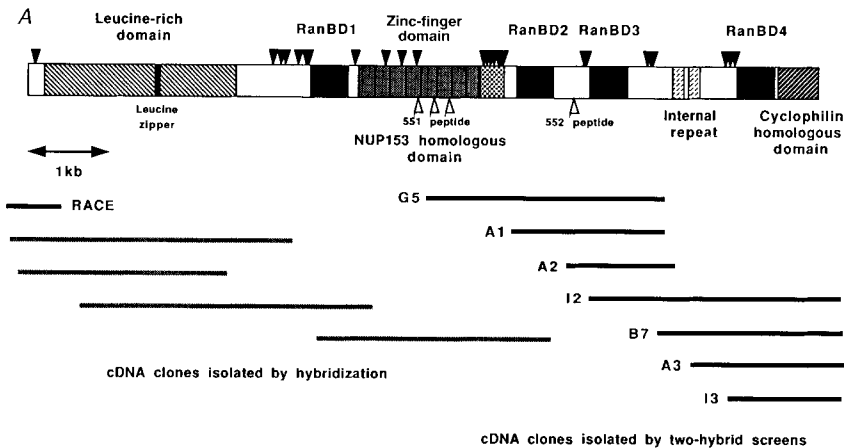
Clones	cDNAs fused to activation domain		Homology	β -galactosidase activity (unit per mg protein)
	Size (kb)	Frequency of cloning		
Vector only				<1
A1	1.9	24	RanBP2	1,395
A2	1.7	6	RanBP2	74
A3	1.8	2	RanBP2	752
B7	2.2	2	RanBP2	580
G5	2.9	2	RanBP2	1,216
I2	3.0	2	RanBP2	66
I3	1.3	2	RanBP2	26
H2	0.9	1	RanBP1	135
B8	1.3	38	Unknown	30
G3	0.8	2	Unknown	28

The plasmids expressing a target protein, Ran/TC4, fused to the GAL4 DNA-binding region, were prepared by inserting the *NcoI*–*Bam*HI fragment of pET–Ran²⁵ into the yeast expression vector pAS1 and introduced into the yeast strain Y190 (ref. 15). The cDNA fused with the GAL4-activation domain of the pACT library, which was prepared from human lymphocytes¹⁵, was introduced into yeast Y190 transformants bearing pAS–RAN, with a LiCl protocol. Transformants were selected on the synthetic medium lacking histidine, tryptophan and leucine in the presence of 25 mM 3-aminotriazole. The isolated colonies were subsequently grown in synthetic medium lacking tryptophan and leucine, and containing 2% ethanol and 3% glycerol, and assayed for β -galactosidase activity. One unit was calculated as 1 nmol o-nitrophenyl- β -D-galactoside cleaved per min per mg protein at 28 °C.

FIG. 1 Cloning and identification of RanBP2. A, Cloning and schematic structure of RanBP2. RACE, a 5' end sequence obtained by the RACE (rapid amplification of cDNA ends) method. Arrowheads indicate the position of XFXFG pentapeptide motifs. RanBD, Ran-binding domain. The open triangles indicate the positions of the peptides used to raise the anti-peptide antibodies 551 and 552. B, Amino-acid sequence of RanBP2. The boxed region indicates the leucine-rich domain, within which the leucine zipper region (residues 450–471) is underlined, with the leucine residues that form the heptad repeat indicated by a heavy underline. The underlined segments outside the box indicate regions that are repeated or significantly homologous with other proteins, such as (a) RanBP1 (refs 9, 10); (b) zinc-finger repeats homologous to NUP153 (refs 11, 12); (c) a region homologous to NUP153 downstream of the zinc-finger domain; (d) an internal repeat; and (e) a cyclophilin¹³-homologous region. The XFXFG pentapeptide motifs are shown by small boxes, and the positions of peptides 551 and 552 used to raise the anti-RanBP2 antibody are highlighted in grey. C, Amino-acid sequences of characteristic domains of RanBP2 were aligned with homologous domains of indicated proteins. Asterisks denote positions occupied by identical amino acids among all the sequences. The (O) symbols show positions that are occupied by the chemically conserved amino acids²⁶ among all the sequences. Ca, Ran/TC4-binding domains; 1–4 are human RanBP2 repeats 1–4; 5 is human RanBP1¹⁰ (residues 12–157). Cb, Zinc-finger domains: 1–8 are human RanBP2 zinc-finger domains 1–8; 9–12 are human NUP153 (ref. 12) zinc-finger domains, 1–4. Cc, Cyclophilin homologous domain: 1, residues 3,064–3,224 of RanBP2; 2, human cyclophilin A¹³.

METHODS. A, Two-hybrid screening was done as described in Table 1. A human cDNA library derived from human B-cell lymphoma cells (CLONTECH) was screened with the 5' end of G5 clone as a probe. The 5' end was determined by the RACE method with the 5' AmpliFINDER RACE kit (CLONTECH). B, The amino acid sequence was deduced from the nucleotide sequence deposited in the DDBJ/EMBL/GenBank database with accession number D42063. C, The percentage identities were statistically calculated by the previously described method²⁷ as follows. The identity between human RanBP1 and the RanBDs, 1, 2, 3 and 4 of human RanBP2: 60% ($P < 10^{-13}$), 49% ($P < 10^{-13}$), 46% ($P < 10^{-13}$), 53% ($P < 10^{-13}$). The cyclophilin domain of RanBP2 and cyclophilin A show a 67% identity ($P < 10^{-13}$).

* To whom correspondence should be addressed.



Ca

1. TDGGSAGHDGDDDDGHPEFVPVLPDKLEVKTGEDEEFFCNRAKLFRFF--DVESKWKERGINVKVILRHKHTSGKIRILL
2. EKDDDAYKTDSDDIHPEFVPVQKPELVLTGEDEEKVYSQVRVLFRFF--DAEVSQWKERGLNGLIKLVENVNGKIRML
3. DEESDVTQEEERDQGFPEFVPVLPDLVEYSGCEENQVVFSHRAKLYRY--DRDVSQWKERIGDGIKILVYNNDKQVRIV
4. DEDSGDESVHNNEDIHEFIPVLSLPE--VEVKSDEGDEDLFPERAKLYRW--KDKVSQWKERGVDGIKILWHTMKNNYRIL
5. DTSTENTDESNDHP--QFEPVLSLPE--QETKLEDEEELFKMRAKLRFASENDLFEWKERGTDGVKLLKHKEGXAIRLL
0 ***0* *0 *0 0 **0* * * *0* 0 0 0 ***** 00*0* * 0
1. MRREQVLKICANHIIISDPMKLTPNAGSDRSFVW--HALDYADELPKPEQLAIRFKPTEPAALFCKKFFEE
2. MRREQVLKVCANHNYITMNLKPLSGSDRAWHW--LASDFSGDGAKEVHLQAAKFTEPLAEFFQKQFEE
3. MRRDQVLIKCANHRIPTDMLKNMKGTERRVWLW--TASDFADGERKVEHLVARFKLQDVAISFFKIFDE
4. MRDQVFKVCANHVIKTMELKPLNVSNNAALVW--TASDYADGAKEVGLAVRFKTEVADCFKTFKEE
5. MRDRDKTLKICANHVIITMMLKPNAGSDRAWWNTHADFADECPPELLAIRFLNAENAQKFKTFKEE
0* *0**0* * * 00 * 00* * * *0* * * *0* * * *0* *

Cb

- [illegible]

Cc

- [illegible]

B

HRRSADKADVERIYASVQSGTSPSPRQSKMQFYFAKLLYEAKYEYDLARKYICTYINVQERDPAKRRPLGLLYELEENTDKAVECYRRSVELNPTQKDLVLKI
 AELLCKNDVTDGRAKYWLERAALFPGSPAIIYKLEQLDCEGEDGWNKFLDILQSELYVRPDDVHVNIIRLVEVYRSTRKRLDAVACHAEARNIALRSS
 LEWNSCVQVTLKEYLESQCLESDKSDWRATNTDLLAYANLMLTLSTRDVQESRELLQSFDSALQSVKSLGGNDELSATFLEMKGPHYMHAGSLLKLM
 GQHSNSVQWRALSELAALCYLIAFQVPRPKIKLIGKEAGQNLLEMMACDRLSQSGHMLLNLNRGQDFLKEIVETFANKSGQSALYDALFSSQPKDTSF
 LGSDDIGNIDVREPELEDLTRYDVGAIRAHNGSLQHLTMLGLQWNSLPALPGIRKWLKQLFHHLPHTSRLTNETAPESICILDEVFLGLVVYTSHLQLK
 ECKNSHSSSYQPLCLPLPVCKQLCTERQKSWMDAVCTLIHRKAVPGNVAKRLRLVQHEINTLRAQEKGHLQALLVHVAECLQKTSGSLNSFYDQREYIG
 RSVHYWKVPLLLKIIKKNSIPEPIDPLFKHFSVDIQASEIVEYEEDAHITFAILDAVNGNIEDAVTAFESIKSVSVYWNLALIFHRKAEDIENDALS
 PEEQECKNYLRKTRDYLIKIIDSDSNLSVVKLPLVPLESVKEMLSVMQLEDEYSEGGLYKNGSLNADSEIKHSTPSPTKYSLSPSKSYKSPKPT
 PRMAEDQNSLLKMICQVEAIKKEMQELKLNSSNSASPHRPTENTGPDSPVDPQYQSGQTFHGAPLTVATGTPSVYISQSPAYNSQYLLRPAANVTPTKG
 PVYGMRLPQQHIIYAYPQQMHTPPVQSSSACMFQSEMYGPPALRFESPATGILSPRGDDYFNYNVQQTSTNPPLPEPGYFTKPIIAHASRSAESKFTB
 FGKTNFVQPMPEGELRSLPTQAHTTQPTFFKFSBNFKSNDQFTFSSPQVVTQPPPAAYSNSSELGLLTSDKPLQGDGYSGAKPIPGGQTIGPRNFTN
 FGKFNVSIGISFTENMGSSQKMSGFRSRDDMTFHBQPCKSYFGPTLETANKNHTDGGSAHGDDDDGPHFEPVPLDPKIEVKTEGEDEKEFFCNRAK
 LFRFDVESKEWKERGIGWVILRHKTSCKIRLLMRREQVLKICANHYSIPDMKLTPNAGSDRSFVHVALDYADELPKPEQLAIRFTPEEALFKCKFEE
 AQSILKAPGTNVAMASNAQVIRVKEPTSHDNKDKICKSDAGNLNFEFQAKKGEWHNHCHSCSLKNASTAKKCVSCQNLMPNSKELVGOPLAETVPTPKTS
 PENVQDRFALVTPKKEGHWDCSILVRNEPTVSRACIAQNTKSAKSGSSFFVHQASSPFQDGDLPKPINSDFRSVFSTKEGQWDCSACLQNEGSSKCA
 ACQNPQRKQSLPATSIPTPASFFKFGTSETSKTLKSGFEDMFAKKEGQWDCSSCLVRNEANATRCVACQNPDKPSPSTSVPAPSFKFGTSETSKAKPSQF
 QMFTTKEGQWDCSCLVRNEASATKCIACQNPQKQNTTSAVTPASSETSKAKPSQFQMGFTTKEGQWDCSCLVRNEASATKCIACQNPQKQNTTSA
 VSTPASSETSKAKPSQFQMGFTTKEGQWDCSCLVRNEASATKCIACQNPQKQNTTSAVTPASSETSKAKPSGFGEMFIRKGQWDCSCLVCCVQNESSSLK
 CVACDASKPTKPIAEAPSAFTLGSEMKLHDSGSSQVGTGFKSNFSEKSKFGTQEGFKFQHVQDQENSSFMNQSSNTEFKSTKEGFSIPVSADGFKF
 DISEPCHQEKSEKPLENGTQFGQAQDISGCKRGRVIFQDTSSTSTTADLAKSTSGEQFGFQKKDPNFKGFSGAGEKLFSSQYQKMANKANTSGDGEKDD
 DAYKTEDSDDIHFEPPVQMPEKVELVTGEDEKVLVSQRVKLFRFDAEVSQWKERGGLNLKILKNEVNGKRLMLMRREQVLKVCANHWITTTNKLPLSG
 SDRAMHMLASDFSDDGAKLEQLAAKFPTPELAEFFKQFEECQRLLLDIPLTQPHKLVDTGRAAKLIQRAEEMKSGLKQFKFTFLTNDQTKVTEENKSGG
 TGAAGASDTTIKPNPENTQFTLEWMDTLREDALDSDSVSSSVBRASPLASSPVKRLFRFGESTIGFNFSFKSALSPSKSPAKLNQSGTSVGTDEESDVT
 QEEERDQYTFEPVPLPDLVEVSSGEENEQVVSRAKLYRYRDVGGQWKEGIGDRIKILQNYDMKQVRIVNRDDQVLKLCANHRITPDMTLQMKGTET
 VHLWTACDFADGERKVEHLAVRFLQDVADSPFKIPDEAKTAQEKDSLITPHVSRSSTPRESPCGKIAVAVLEETTRERTDVIQGDVADATSEVEVSST
 SETTPKAVVSPPLFVFGSESVKSIFSSSEKSKFFAFGNSATSGSLFQFSFNAPLKSNNSETSSVAQSGSESKEVPKKELSKNSDIEQSSDSVKVNLFAF
 PTEESSINYFTKTPKAKAKKKPEDSPDDDLIVYELTPTAEQKALATKLPLPTTFYCKNRPDVYSEEEDEDDFETAVUKLNGKLYLDGSEKCRPLE
 ENTADNEKECIIVMEKKPTVEEKAKADTLKLPLPTTFYCGVCSDDTEDNGNDEQFSELQKVQEAQKSQTEETSTSDSVYTGTEVMVSPFCSEEPDSIT
 KSISPPSSSETMDKPDVLSRKEIDTSTSGGESKIVSFGFGSSTSGLSFADLASSNSGDFAFGSKDKMFWANTQRAVFGQSVGTQSAGKVGEDGGS
 DEEVVHNEIDHFEPIVLSPEVEVKSGEEDDEILFKERAKLYRMDRVSQWKERGVGDIKILWHTMKNYRIYLMRRDQVFKVCANHWITTKTMELKPLVSN
 NALVMTASDYADGEAKVEQLAVRFLKTEVADCFKKTFEECQQLMLKLGHVLSAAELSKETNPVVFFDVCADGEPLGRITMELFSNIVPRTAEFRALC
 TGEKGFQFKNSIFHRVIPDPVCGGDDITKHDGTGGQSIYGDKFEDENFDVKHTGPGLLSHANQCGQNTNNSQFVITLTKAEHLDFKHVVFQVVDGMDTVK

near the 5' end of the cloned cDNA. The nucleotide sequence (GGCGCGATG) at the putative N-terminal end of the coding sequence corresponds to the Kozak consensus¹⁶, and there are no ATG codons before it. Although the absence of in-frame termination codons upstream allows the possibility that a small extension at the N terminus remains to be found, the putative ORF encodes a predicted protein of relative molecular mass (M_r) 358 K, in good agreement with the size of a polypeptide recognized by immunoblotting with affinity-purified antibodies raised against RanBP2 (Fig. 2b, lanes 1–3).

Consistent with the isolation of RanBP2 using Ran/TC4 as a 'bait', Ran/TC4-GTP (and not Ran/TC4-GDP) bound RanBP2 as detected by immunoprecipitation with an anti-RanBP2 antibody (Fig. 2c) and the sequence of RanBP2 showed four domains with strong homology to RanBP1 (refs 9, 10; Fig. 1C). To test for their ability to bind Ran/TC4-GTP, each of

these domains was separately expressed in *Escherichia coli*, run on SDS-PAGE and analysed by overlay blotting with Ran/TC4 loaded with [³⁵S]GTP-γS and [³⁵S]GDP-βS (Fig. 2c). All domains specifically bind Ran/TC4-GTP; they are therefore designated as Ran-binding domains (RanBDs) by functional as well as by sequence similarity criteria. Although the affinity for binding Ran/TC4-GTP is diverse among domains, these RanBDs may function cooperatively *in vivo*. RanBP2 possesses 26 copies of a degenerate XFXFG pentapeptide motif that is a diagnostic feature of a series of NPC proteins¹⁴ (Fig. 1). RanBP2 co-fractionates with rat-liver nuclear envelopes (Fig. 2b, lanes 4–6) and recognized with an anti-NPC monoclonal antibody, MAb350 (refs 11, 17) (Fig. 2b, lanes 9 and 13). It is the largest protein immunoprecipitated with the MAb350. By confocal microscopy of digitonin-permeabilized HeLa cells, the anti-551 peptide RanBP2 antibody revealed a punctuated nuclear surface

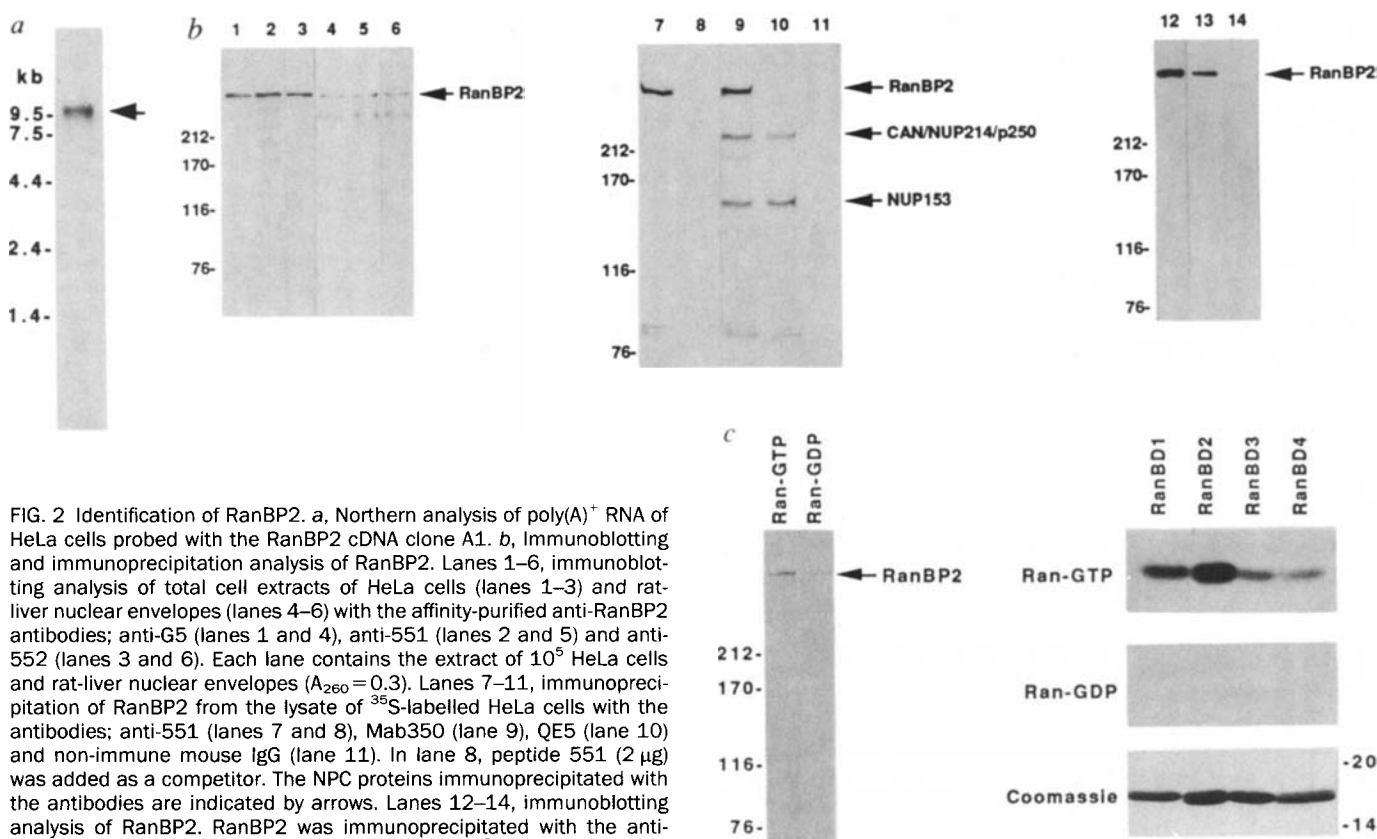


FIG. 2 Identification of RanBP2. **a**, Northern analysis of poly(A)⁺ RNA of HeLa cells probed with the RanBP2 cDNA clone A1. **b**, Immunoblotting and immunoprecipitation analysis of RanBP2. Lanes 1–6, immunoblotting analysis of total cell extracts of HeLa cells (lanes 1–3) and rat-liver nuclear envelopes (lanes 4–6) with the affinity-purified anti-RanBP2 antibodies; anti-G5 (lanes 1 and 4), anti-551 (lanes 2 and 5) and anti-552 (lanes 3 and 6). Each lane contains the extract of 10^5 HeLa cells and rat-liver nuclear envelopes ($A_{260} = 0.3$). Lanes 7–11, immunoprecipitation of RanBP2 from the lysate of ³⁵S-labelled HeLa cells with the antibodies; anti-551 (lanes 7 and 8), Mab350 (lane 9), QE5 (lane 10) and non-immune mouse IgG (lane 11). In lane 8, peptide 551 (2 μg) was added as a competitor. The NPC proteins immunoprecipitated with the antibodies are indicated by arrows. Lanes 12–14, immunoblotting analysis of RanBP2. RanBP2 was immunoprecipitated with the anti-551 antibody from the lysate of HeLa cells (3×10^6 cells), analysed by 5% SDS-PAGE and immunoblotted using the antibodies as a probe: anti-551 (lane 12), Mab350 (lane 13) and QE5 (lane 14). **c**, Ran/TC4-GTP, but not Ran/TC4-GDP, binds to RanBP2 and its Ran-binding domains (RanBD). Left panel, autoradiograph of a blot of the immunoprecipitated RanBP2 with [³⁵S]GTP-γS- and [³⁵S]GDP-βS-Ran/TC4. Right panel, autoradiograph of a blot of *E. coli*-produced RanBDs 1–4 with [³⁵S]GTP-γS- and [³⁵S]GDP-βS-Ran/TC4.

METHODS. **a**, Northern analysis was carried out as described²⁸. RNA marker sizes are shown in kilobases. **b**, Lanes 1–6: Immunoblotting analyses were carried out as described^{3,28}. We prepared three kinds of antibodies to RanBP2, against peptides 551 and 552 (see Fig. 1), and using *E. coli*-produced GST-fused G5 protein which were affinity-purified by using antigen-coupled Sepharose columns. HeLa cells were lysed as described¹⁸. Rat liver nuclear envelope was prepared as described²⁹, which was also used for immuno-electron microscopy. Samples were analysed by 5% SDS-PAGE and immunoblotted using indicated antibodies as described^{3,28}. Lanes 7–11: HeLa cells labelled for 6 h with [³⁵S]methionine (200 μCi ml⁻¹) were lysed on ice in the lysis buffer as described¹⁸. After centrifugation, the supernatant was mixed with the indicated antibodies. For the immunoprecipitation with QE5 and Mab350, 15 μg of anti-mouse rabbit antibody was used as bridging antibody. The immuno-complexes were analysed by 5% SDS-PAGE, and exposed to X-ray film. Lanes 12–14: RanBP2 immunoprecipitated with

the anti-551 antibody was analysed by 5% SDS-PAGE and immunoblotted using the indicated antibodies as probes. Marker sizes ($10^{-3} \times M_r$) are shown. **c**, Blots of RanBP2 were prepared as described in **b**, lanes 12–14. The Ran-binding domains of RanBP2 (RanBD 1 (residues 1,155–1,321), RanBD2 (residues 1,996–2,153), RanBD3 (2,293–2,446) and RanBD4 (2,895–3,045)) were subcloned into pQE30 or pQE31 vectors (Qiagen), and expressed in *E. coli* XL1-blue. Recombinant RanBDs were purified using Ni-NTA-agarose resin (Qiagen), analysed by 15% SDS-PAGE and transferred to nitrocellulose filters. Blots were renatured at 4 °C overnight as described³⁰, incubated for 3 h in PBS containing 0.3% Tween-20, 5 mM dithiothreitol, preincubated for 30 min at room temperature in a buffer containing 0.05% Tween-20, 2 mM dithiothreitol and 50 μM unlabelled GTP in PBS, and then overlaid with [³⁵S]GTP-γS- or [³⁵S]GDP-βS-Ran/TC4 for 30 min at room temperature. After washing with binding buffer, blots were dried and exposed to X-ray film. [³⁵S]GTP-γS- and [³⁵S]GDP-βS-Ran/TC4 were generated by loading 75 μg of recombinant Ran/TC4 with 2 nmol of unlabelled GTP or GDP, and 50 μCi [³⁵S]GTP-γS or [³⁵S]GDP-βS for 45 min on ice in the presence of 20 mM HEPES, pH 7.6, 1 mM MgCl₂, 10 mM EDTA, 2 mM dithiothreitol in a 50-μl reaction volume. The reaction was stopped by addition of 20 mM (final concentration) MgCl₂ and unincorporated nucleotides, were removed by gel filtration. Marker sizes ($10^{-3} \times M_r$) are shown.

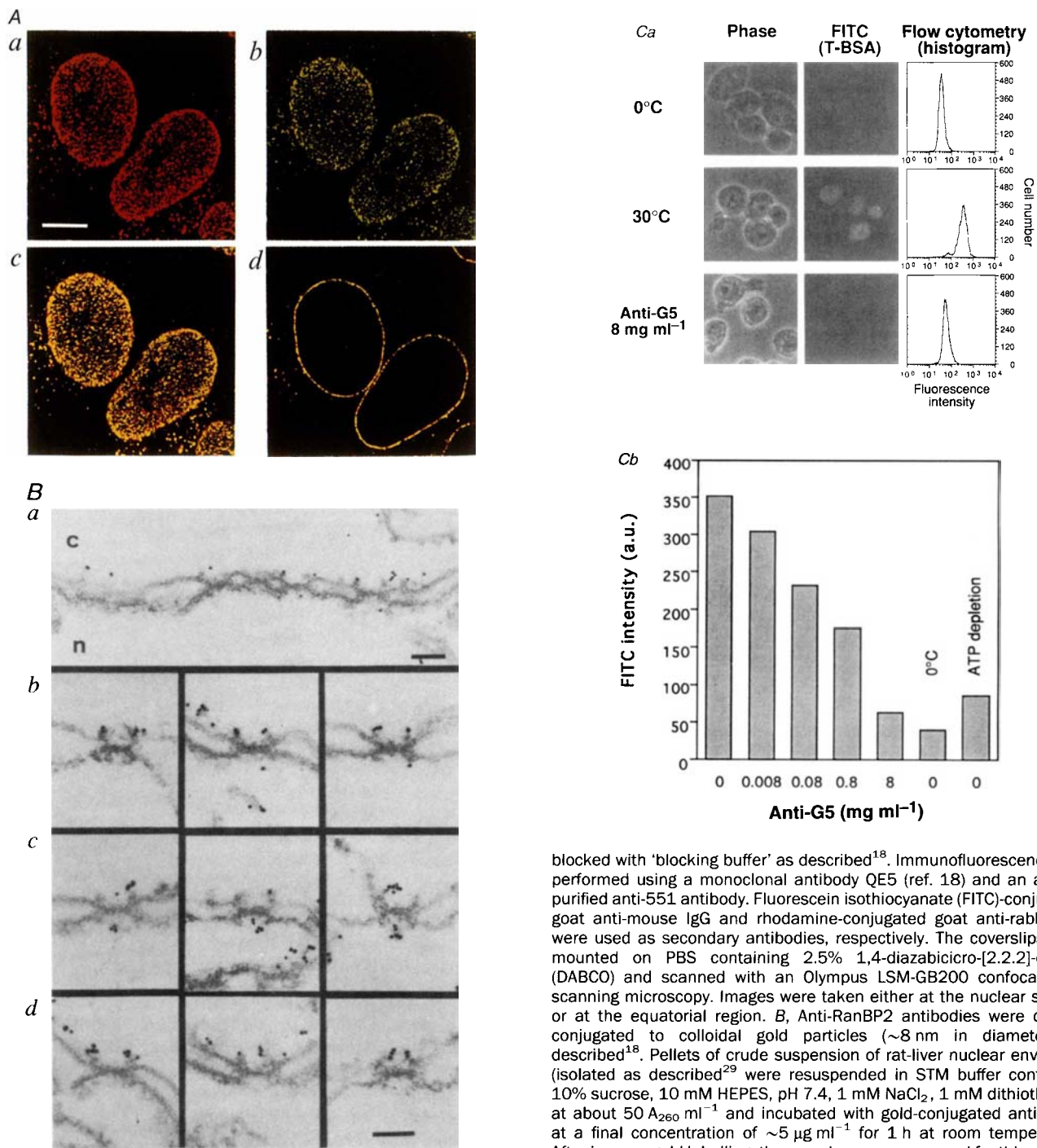


FIG. 3 RanBP2 is located in the NPC and is required for nuclear import. **A**, Colocalization of RanBP2 with NPC in HeLa cells by immunofluorescence staining: **Aa**, Stained by the anti-551 antibody and visualized at the nuclear surface; **Ab**, stained by the QE5 antibody and visualized at the nuclear surface; **Ac**, superimposition of these two images by electronic image processing; **Ad**, superimposition of views of **a** and **b** taken at the equatorial plane. Bar, 10 μ m. **B**, Localization of RanBP2 in isolated rat-liver nuclear envelopes by immunogold electron microscopy: **Ba**, view along a single nuclear envelope labelled with the anti-G5 antibody; **Bb–d**, gallery of selected NPCs labelled with the anti-RanBP2 antibodies, anti-G5 (**b**), anti-551 (**c**) and anti-552 (**d**). The cytoplasmic (**c**) and nuclear (**n**) surface of the nuclear envelope are indicated in **a**. Scale bars, 100 nm. **C**, Nuclear import assay: **Ca**, nuclear transport of NLS-coupled BSA in the presence of the anti-G5 RanBP2 antibody; **Cb**, dose-dependent inhibition of nuclear transport with the anti-G5 RanBP2 antibody.

METHODS. **A**, HeLa cells growing on coverslips were fixed with 4% paraformaldehyde, permeabilized with 40 μ g ml⁻¹ digitonin, and

blocked with 'blocking buffer' as described¹⁸. Immunofluorescence was performed using a monoclonal antibody QE5 (ref. 18) and an affinity-purified anti-551 antibody. Fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG and rhodamine-conjugated goat anti-rabbit IgG were used as secondary antibodies, respectively. The coverslips were mounted on PBS containing 2.5% 1,4-diazabicyclo-[2.2.2]-octane (DABCO) and scanned with an Olympus LSM-GB200 confocal laser scanning microscopy. Images were taken either at the nuclear surface or at the equatorial region. **B**, Anti-RanBP2 antibodies were directly conjugated to colloidal gold particles (~8 nm in diameter) as described¹⁸. Pellets of crude suspension of rat-liver nuclear envelopes (isolated as described²⁹ were resuspended in STM buffer containing 10% sucrose, 10 mM HEPES, pH 7.4, 1 mM NaCl₂, 1 mM dithiothreitol, at about 50 A₂₆₀ ml⁻¹ and incubated with gold-conjugated antibodies at a final concentration of ~5 μ g ml⁻¹ for 1 h at room temperature. After immunogold labelling, the samples were processed for thin-section EM by conventional procedures as described³¹. **C**, HeLa cells were permeabilized with 80 μ g ml⁻¹ digitonin in 'transport buffer' as described¹⁹. Various concentrations of the affinity-purified anti-G5 antibody were added to 20 μ l of reaction mixtures containing 10⁵ permeabilized HeLa cells and 4 mg ml⁻¹ HeLa-cell cytosolic extract and the substrate for nuclear transport (10 μ g ml⁻¹ FITC-labelled BSA coupled to a peptide containing the NLS of SV40 T antigen (T-BSA)⁹ in 'transport buffer' containing 10 mg ml⁻¹ BSA and 1 mM GTP. The final concentration of total IgG was adjusted to 8 mg ml⁻¹ for each reaction mixture, by mixing rabbit non-specific IgG with the anti-G5 antibody. ATP was depleted by the addition of 20 U hexokinase and 5 mM glucose. After incubation either at 30 °C or 0 °C for 30 min, reactions were stopped by the addition of 2 ml of the ice-cold transport buffer containing 10 mg ml⁻¹ BSA, and then fixed with 4% paraformaldehyde. For microscopic analysis, fixed cells were mounted on PBS containing 2.5% DABCO and then photographed with a Zeiss Axiophot fluorescence microscope. To determine the mean fluorescence intensity of FITC per cell, fixed cells were suspended in 1 ml of PBS and analysed with a Becton Dickinson FACScan flow cytometer³². Values are in arbitrary units (a.u.).

and ring staining (Fig. 3A) that largely colocalized with the staining of QE5 (ref. 18), which recognizes CAN/NUP214/p250, NUP153 and p62, but not RanBP2 (Fig. 2b, lanes 10 and 14). Because digitonin permeabilizes only the plasma membrane¹⁹, this finding indicates that a portion of RanBP2 is located at the cytoplasmic face of the NPC. To examine the location of RanBP2 in more detail, we performed immunogold electron microscopy on isolated rat liver nuclear envelopes with gold-conjugated antibodies raised against three regions of RanBP2 indicated in Fig. 1A, B. All antibodies labelled the NPC at the cytoplasmic fibrils emanating from the cytoplasmic ring of the NPC (Fig. 3B). In addition, the anti-551 peptide and the anti-G5, but not the anti-552 peptide, antibodies labelled the nuclear face of the NPC. As both anti-551 and G5 antibodies do not seem to crossreact with other NPC proteins in western blots (Fig. 2b), it is possible that RanBP2 is localized on both sides of the NPC and the epitopes of the 552 antibody are not exposed at the nuclear side. Another possibility is that both the anti-551 and G5 antibodies that were raised against the region of RanBP2 homologous to NUP153 (see Fig. 1) recognize NUP153 in its native conformation. We tested the ability of the anti-G5 antibody to inhibit nuclear import of an NLS-fused protein in a dose-dependent manner (Fig. 3C). Addition of 8 mg ml⁻¹ of the antibody inhibited nuclear import almost completely, whereas no inhibition was observed on addition of 8 mg ml⁻¹ of the control IgG. The low-molecular-mass (~20K) dextran normally diffused into the nuclei in the presence of anti-G5 antibody (results not shown). Thus our data indicate that RanBP2 is an NPC protein and specifically binds Ran/TC4-GTP, which has been shown to be essential for importing NLS-bearing proteins^{20,21}. In agreement with our findings, Melchior and co-workers showed that Ran/TC4-GTP binds to the cytoplasmic surface of the NPC, and, using our anti-RanBP2 antibody:anti-551, identified RanBP2 as the only prominent protein of rat-liver nuclear envelopes that binds Ran-GTP (F. Melchior, T. Guan, N.Y., T.N. and L. Gerace, submitted).

RanBP2 has eight zinc-finger motifs, which is same as that of NUP153 (refs 11, 12) (Fig. 1Cb). Similarly to NUP153 (ref. 11), the zinc-finger domains of RanBP2 bound labelled DNA in a Zn²⁺-dependent manner (results not shown). The functional significance of these zinc-finger motifs remain elusive. The presence of the same zinc-finger motif in both the cytoplasmic fibrils and the nuclear baskets, where NUP153 is located¹⁸, suggests its involvement in a common mechanism for nuclear import and export.

At its C terminus, RanBP2 possesses a domain that shows very high homology with cyclophilin¹³ (Fig. 1Cc). Although cyclophilin is well conserved through evolution, and can act as a peptidyl-prolyl isomerase¹³, its precise function is unknown. The N-terminal part (residues 66–768) of RanBP2 consists of a leucine-rich domain containing a leucine zipper motif (Fig. 1B) like the other NPC proteins, CAN/NUP214/p250 (refs 22, 23) and NUP107 (ref. 24). It will be very interesting to learn what kind of protein binds the N-terminal leucine-rich domain of RanBP2. □

12. McMorow, I., Bastos, R., Horton, H. & Burke, B. *Biochim. biophys. Acta* **1217**, 219–223 (1994).
13. Galat, A. *Eur. J. Biochem.* **216**, 689–707 (1993).
14. Panté, N. & Aebi, U. *J. struct. Biol.* **113**, 179–190 (1994).
15. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75**, 805–816 (1993).
16. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
17. Davis, L. I. & Fink, G. R. *Cell* **61**, 965–978 (1990).
18. Panté, N., Bastos, R., McMorow, I., Burke, B. & Aebi, U. *J. Cell Biol.* **126**, 603–617 (1994).
19. Adam, S. A., Sterne-Marr, R. E. & Gerace, L. *J. Cell Biol.* **111**, 807–816 (1990).
20. Moore, M. S. & Blobel, G. *Nature* **365**, 661–663 (1993).
21. Melchior, F., Paschal, B., Evans, J. & Gerace, L. *J. Cell Biol.* **123**, 1649–1659 (1993).
22. Von Lindern, et al. *Molec. cell. Biol.* **12**, 1687–1697 (1992).
23. Kraemer, D., Wozniak, R. W., Blobel, G. & Radu, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1519–1523 (1994).
24. Radu, A., Blobel, G. & Wozniak, R. W. *J. biol. Chem.* **269**, 17600–17605 (1994).
25. Dasso, M., Seki, T., Azuma, Y., Ohba, T. & Nishimoto, T. *EMBO J.* **13**, 5732–5744 (1994).
26. Schwartz, R. M. & Dayhoff, M. O. in *Atlas of protein sequence and structure*, Vol. 5 suppl. 3 (ed. Dayhoff, M. O.) 353–358 (National Biomedical Research Foundation, Washington DC, 1978).
27. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827–829 (1983).
28. Ohtsubo, M. et al. *EMBO J.* **10**, 1265–1273 (1991).
29. Gerace, L., Comeau, C. & Benson, M. *J. Cell Sci. Suppl.* **1**, 137–160 (1984).
30. Lounsbury, K. M., Beddow, A. L. & Macara, I. G. *J. biol. Chem.* **269**, 11285–11290 (1994).
31. Byrd, D. A. et al. *J. Cell Biol.* **127**, 1515–1526 (1994).
32. Paschal, B. M. & Gerace, L. *J. Cell Biol.* **129**, 925–937 (1995).

ACKNOWLEDGEMENTS. We thank S. J. Elledge for the cDNA libraries, plasmids and yeast strains; K. Chida, T. Tachibana and Y. Yoneda for the proteins fused with NLS; the Human Genomic Center for supplying the sequence data; F. Melchior and L. Gerace for informing us of their preliminary results; B. Paschal for providing, before publication, the protocol for analysing nuclear transport by flow cytometry; L. I. Davis and B. Burke for providing the MAb350 and the QE5 antibody; U. Sauder for help with embedding and sectioning; and T. Hunt for critical reading of the manuscript. This work was supported by Grants-in-Aid for Scientific Research and for Cancer Research from the Ministry of Education, Science and Culture of Japan, and by the HFSP.

Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways

Bing Xiao*, Stephen J. Smerdon*, David H. Jones*, Guy G. Dodson*†, Yasmina Soneji*, Alastair Aitken* & Steven J. Gamblin*

* Division of Protein Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† Department of Chemistry, University of York, Heslington, York YO1 5DD, UK

A BROAD range of organisms and tissues contain 14-3-3 proteins, which have been associated with many diverse functions including critical roles in signal transduction pathways, exocytosis and cell cycle regulation¹. We report here the crystal structure of the human T-cell 14-3-3 isoform (τ) dimer at 2.6 Å resolution. Each monomer (M , 28K) is composed of an unusual arrangement of nine antiparallel α -helices organized as two structural domains. The dimer creates a large, negatively charged channel approximately 35 Å broad, 35 Å wide and 20 Å deep. Overall, invariant residues line the interior of this channel whereas the more variable residues are distributed on the outer surface. At the base of this channel is a 16-residue segment of 14-3-3 which has been implicated in the binding of 14-3-3 to protein kinase C.

The 14-3-3 family of proteins (named according to their electrophoretic mobility²) contains five major mammalian forms and at least another ten from yeast, plants, amphibians and invertebrates¹. All members of the family show a high sequence similarity. Details of the crystal structure determination of the 14-3-3 τ isoform are summarized in Table 1. The initial electron density map was of excellent quality, and most of the side chains were readily built. The present model accounts for 436 residues of the dimer and 107 solvent molecules. The τ dimer contains six motifs of three-helix bundles (Figs 1a, b and 2a). The larger amino-terminal domain provides both the dimer interface and

Received 10 November 1994; accepted 9 June 1995.

1. Dasso, M. *Trends biochem. Sci.* **18**, 96–101 (1993).
2. Bischoff, F. R., Maier, G., Tilz, G. & Ponstingl, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8617–8621 (1990).
3. Nishitani, H. et al. *EMBO J.* **10**, 1555–1564 (1991).
4. Sazer, S. & Nurse, P. *EMBO J.* **13**, 606–615 (1994).
5. Kadowaki, T., Goldfarb, D., Spitz, L. M., Tartakoff, A. M. & Ohno, M. *EMBO J.* **12**, 2929–2937 (1993).
6. Cheng, Y., Dahlberg, J. E. & Lund, E. *Science* **267**, 1807–1810 (1995).
7. Amberg, D. C., Fleischmann, M., Stagljar, I., Cole, N. & Aebi, M. *EMBO J.* **12**, 233–241 (1993).
8. Tachibana, T., Imamoto, N., Seino, H., Nishimoto, T. & Yoneda, Y. *J. biol. Chem.* **269**, 24542–24545 (1994).
9. Coutavas, E., Ren, M., Oppenheim, J. D., D'Eustachio, P. & Rush, M. G. *Nature* **366**, 585–587 (1993).
10. Hayashi, N. et al. *Molec. gen. Genet.* (in the press).
11. Sukegawa, J. & Blobel, G. *Cell* **72**, 29–38 (1993).

TABLE 1 Structure determination

Data set	No. of crystals	Resolution (Å)	Unique reflections	complete %	R_{sym}	R_{iso}	Phasing power	R_c
Native	1	2.6	17,169	99.7	8.9			
MMC	1	3.4	7,923	97.0	14.1	27.6	1.55	0.67
PTC	1	3.9	5,644	95.2	9.7	22.5	0.95	0.78
Phasing statistics (15–3.4 Å)								
Resolution (Å)	9.80	6.80	5.40	4.40	3.80	3.40		
FOM	0.85	0.63	0.61	0.51	0.34	0.21		
Refinement (15–2.6 Å)								
		(work set)	(free set)					
No. of reflections		16,351	818					
No. of protein atoms		3,461						
No. of solvent molecules		107						
R_{factor} (%)		22.4	28.7					
r.m.s. bond length (Å)		0.013						
r.m.s. bond angles (deg)		1.4						

Full details of the structure determination will be presented elsewhere. The τ isoform of 14-3-3 protein (expressed in *E. coli*¹⁵) was crystallized from 24% PEG 4K, 140 mM Li_2SO_4 , 4% of saturated NaCl in 80 mM MES at pH 6.4 (protein concentration 7 mg ml⁻¹) by vapour diffusion. The crystals belong to space group $P2_12_12_1$ with cell dimensions $a=70.29$ Å, $b=79.30$ Å and $c=101.00$ Å. The needle-like crystals appear over 7 to 10 days and reach maximum dimensions of 0.005 mm $\times$ 0.009 mm $\times$ 0.200 mm. The crystallographic asymmetric unit contains two τ monomers (M_r 28K) and approximately 50% solvent. Crystals were mounted in rayon loops¹⁶ and maintained at a temperature of 100 K using an Oxford cryosystem. The data were collected at SRS Daresbury 9.6 and BL4/ID2 at ESRF Grenoble, both at around 0.87 Å wavelength using a Marresearch image plate. Crystals were prepared for cryocooling by soaking in crystallization buffer with 20% PEG400. The derivatives were similarly treated with cryobuffer made up to 0.5 and 1.0 mM methyl mercury chloride (MMC) and platinum terpyridine chloride (PTC), respectively, for 5 to 6 h. Data were processed with DENZO and SCALEPACK¹⁷. The four Hg sites were found using HASSP¹⁸ and were refined using MLPHARE¹⁹; these SIR phases were then used to determine the Pt sites by difference Fourier procedure. The MIR map then calculated at 3.8 Å resolution was averaged once according to the local two-fold axis (Hg sites and self-rotation function) using a large spherical mask. This map showed extended regions of connected density and an easily recognized protein solvent boundary, enabling the construction of a rather accurate envelope using ENVATOM²⁰. Initial MIR phases to 3.4 Å were then refined and extended using two-fold non-crystallographic symmetry averaging, solvent flattening and histogram matching together with refinement of the non-crystallographic symmetry operators using DM¹⁸. After several rounds of this procedure, with improved masks, the real space correlation coefficient was 0.92 (15 to 2.6 Å). The maps were of very high quality and the model was built readily using O²¹. XPLOR refinement²² using highly weighted non-strict non-crystallographic symmetry restraints, a solvent mask and all the observed data from 15 to 2.6 Å produced a model with 91% of the residues in the most favoured region of the Ramachandran plot. The present model is missing 4 residues from the N terminus and 12 from the C terminus, and 9 residues have not been placed in the C–D loop. Helices C and D are connected in the display figures between residues 66 and 76 for clarity. Residues are numbered so that the first Met in the sequence is residue number 2 in the structure file. This convention was adopted to facilitate future comparison with other mammalian 14-3-3 proteins, some of which have an extra residue at the N terminus. Full coordinates will be deposited with the Protein Databank in due course. $R_{\text{sym}} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$, where I is the observed intensity, $\langle I \rangle$ is the average intensity from multiple observations of symmetry related reflections. $R_{\text{iso}} = \sum_{hkl} |F_{\text{der}} - F_{\text{nat}}| / \sum_{hkl} F_{\text{nat}}$, where F_{nat} is the protein structure factor amplitude, F_{der} is the heavy-atom derivative structure factor amplitude. $R_c = (\text{lack of closure}) / (\text{isomorphous difference})$, where $\text{lack of closure} = \sum |F_{\text{der}} \pm F_{\text{nat}}| - F_H$, and F_H is the calculated heavy-atom structure factor amplitude. Phasing power is defined as $(F_H / \text{lack of closure})$. Crystallographic $R_{\text{factor}} = \sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}| / \sum_{hkl} F_{\text{obs}}$. The r.m.s. bond length and angles are the deviations from ideal values.

the floor of the channel, whereas the sides or walls of the channel are created by the carboxy-terminal domains. The buried interface between the two monomers is largely hydrophobic, indeed non-polar interactions contribute 1,520 Å² to the total 2,150 Å² of inaccessible surface (Fig. 1a).

Three points emerge from the extent and distribution of the sequence similarity between the various 14-3-3 isoforms (Fig. 1b). Variable residues are generally located on the surface of the protein and are not involved in packing of the α -helices. Thus the τ structure is an excellent generic model for the different members of the 14-3-3 family. Second, the fact that most of the residues that line the inside of the channel are invariant suggests that this surface must recognize common features of target binding proteins. However, a full understanding of the sequence conservation will require the structure determination of 14-3-3 complexes. Third, the largely conserved nature of the residues making up the dimer interface means that heterodimer formation is quite possible in tissues where more than one isoform is expressed.

Since the initial identification of 14-3-3 proteins as activators of tyrosine and tryptophan hydroxylases³, the biological roles attributed to these proteins have increased substantially (refs 1, 4 and references therein). It is emerging that 14-3-3 proteins are important in signal transduction pathways and the cell cycle. The mechanism of 14-3-3 proteins frequently appears to involve

interaction with protein kinases^{1,4}. Three kinases, Raf-1, Bcr-kinase and protein kinase C (PKC) in particular have been shown to be regulated by 14-3-3 proteins^{5,10}.

Three main lines of evidence indicate that kinases may bind to the channel of 14-3-3 (these arguments also apply, in part, to many of the other proteins with which 14-3-3 interacts¹). First, the electrostatic potential surface (Fig. 1c, d) shows that the predominantly invariant channel lining is very acidic, so it is interesting that binding to 14-3-3 has been suggested to involve a rather basic zinc-finger motif, at least for PKC (ref. 2) and Raf-1 (ref. 1). Second, residues 123 to 138 of 14-3-3 are probably involved in binding to PKC^{11,13}. In the structure of 14-3-3 this annexin-like sequence is located in the latter part of helix E extending into the first part of the E–F loop, and defines one corner of the channel (Figs 1a, 2b and 3). Crystal structures are known for annexin I and V, and their two C-terminal domains are quite similar. The last 14 residues of these annexins are folded as a short helix leading into a turn, and superimpose well with the corresponding region of 14-3-3. Although the arrangement of other secondary structure elements surrounding these residues is different between the annexins and 14-3-3, the pattern of exposed residues in both cases identifies candidate binding interactions. The role of these and additional residues from the surrounding protein in the binding of 14-3-3 to protein kinases (and other target proteins) can now be investigated more directly by

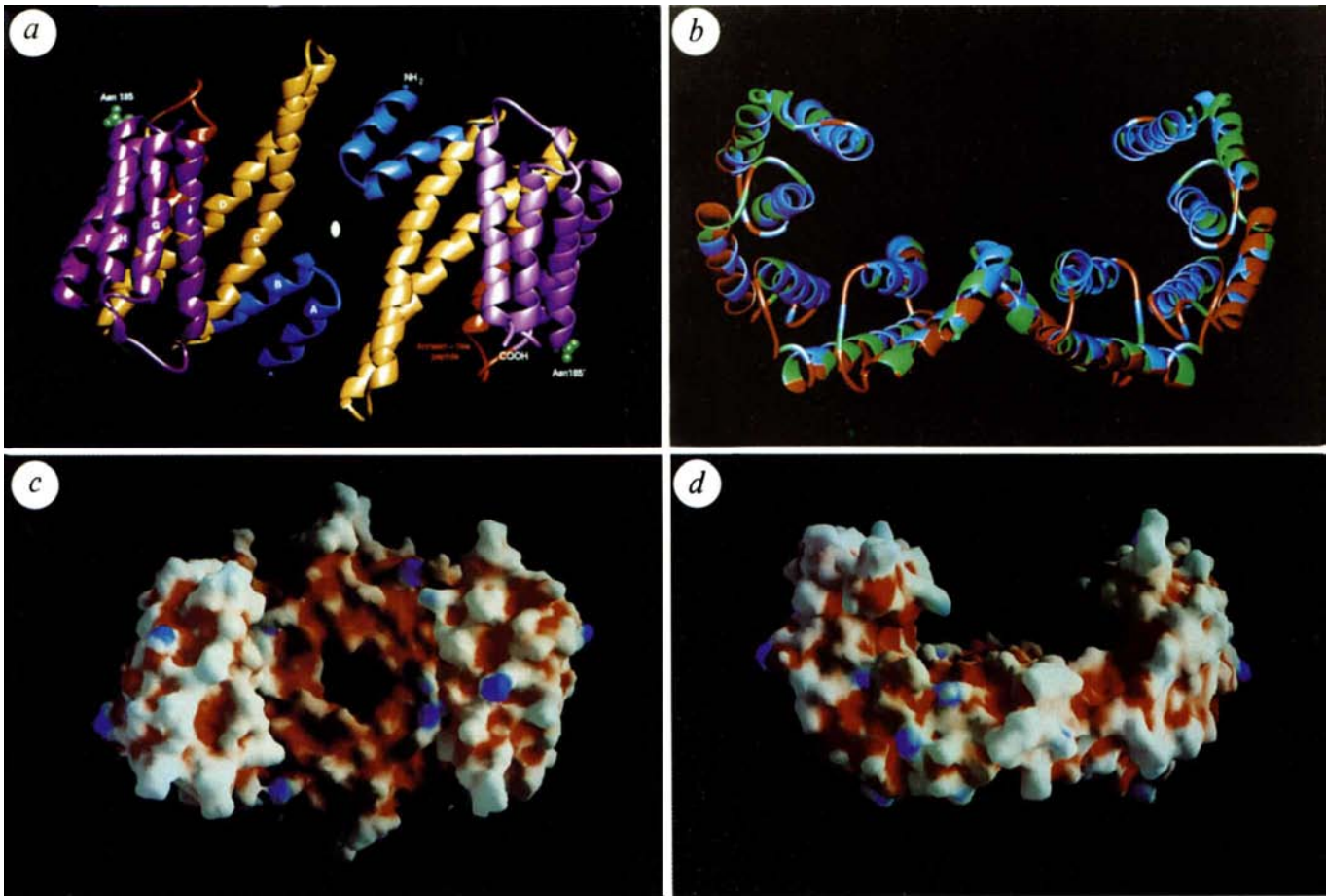


FIG. 1 a, A RIBBONS representation²³ of the τ dimer viewed down the local two-fold axis (shown). The N-terminal domain (which was previously identified by limited proteolysis¹⁵ has helices A and B at the interface (blue) together with helices C and D (gold). The C-terminal domain (helices E to I) are purple. A region of helix E and the E/F loop which shows sequence similarity to the C terminus of the annexin family¹¹ is shown in red. Residue 185 (green) is an asparagine in τ , but is a serine in β and ζ , and a site of *in vivo* phosphorylation in these isoforms. b, Orthogonal view to a of the τ dimer in which the sequence

homology between 24 members of the 14-3-3 superfamily encompassing mammalian, yeast, plant and amphibians isoforms has been mapped onto the crystal structure. Light-blue sequences indicate absolute sequence identity in all 24 proteins, whereas the green and red regions indicate areas of increasing sequence variation. The inner surface of a large channel formed between the C-terminal domains is lined with most of the invariant residues. c, d, Orthogonal views of the electrostatic potential surface of τ generated using the program GRASP²⁴. Negative potential is coloured red and positive potential blue.

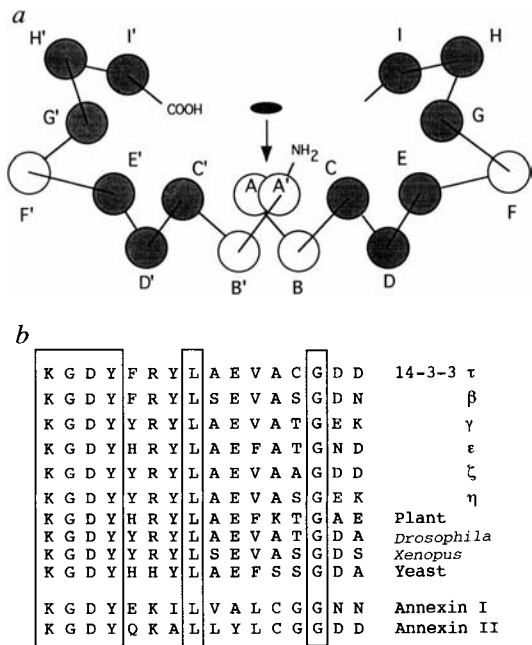
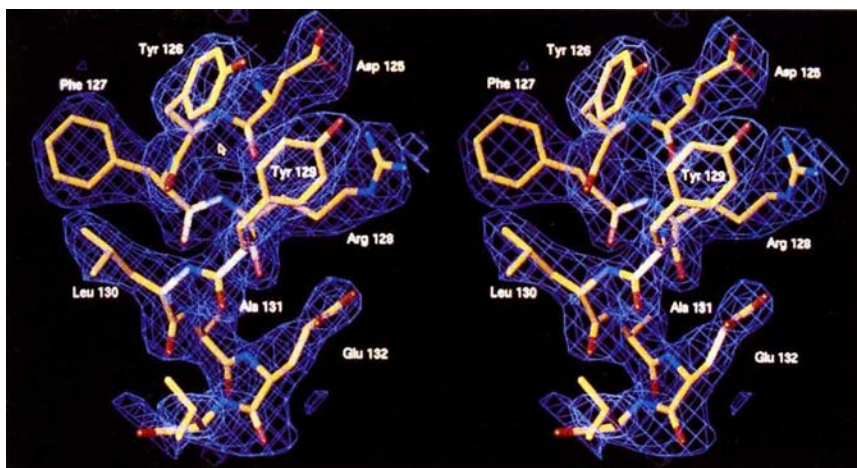


FIG. 2 a, Schematic representation of the τ dimer architecture. The topological arrangement of the nine anti-parallel α -helices of each monomer is shown with the two three-helix bundles of each monomer shaded. Three-helix bundles are also formed across the dimer interface by A/B/C' and its dyad related equivalent A'/B'/C. The approximate position of the dyad axis is indicated by the arrow. b, Sequence similarity between members of the mammalian, amphibian, yeast, insect and plant 14-3-3 family (residues 123 to 138; see legend to Table 1 for numbering convention) and the C-terminal peptides from annexins I (residues 305 to 319) and II (residues 323 to 338)¹¹. A 14-3-3 consensus peptide was found to compete with full-length 14-3-3 for binding to PKC (A.A. and Y.S., unpublished observations). A peptide derived from the C terminus of annexin I was also shown to inhibit the binding of PKC to a number of receptors for activated C kinase¹². In a third study, a peptide derived from the C terminus of annexin II blocked the 14-3-3 stimulation of Ca^{2+} -dependent exocytosis in permeabilized adrenal chromaffin cells¹³.

FIG. 3 A stereo representation of a portion of the electron density map ($2F_o - F_c$) covering residues 125 to 132 (E-helix) contoured at 1.2σ . This section constitutes part of the annexin-like sequence of τ which has been implicated in the binding of 14-3-3 to PKC. There is an extensive network of hydrogen bonds and salt bridges involving Asp 125, Arg 128, Glu 132 (within this helix) and Tyr 149 (not shown).



mutagenesis experiments. Finally, residue 185, which is located in the G-H loop overlooking the annexin-like sequence, is an important *in vivo* phosphorylation site in the β and ζ isoforms¹⁴, which interact with Raf-1 kinase⁴. This phosphorylation site is within a consensus sequence motif for cyclin-dependent kinases¹⁴, and its location lends support to the notion that this channel region constitutes the binding site for kinases.

Because 14-3-3 binds to many kinases it may provide a mechanism by which multiple signal transduction pathways can be simultaneously coordinated. It is also possible that 14-3-3 proteins act independently in signal transduction pathways as mediators, allowing various kinases to associate in a controlled manner⁴. Given the relatively large channel between the two proposed binding sites it is possible that two kinases could bind simultaneously to 14-3-3. One of the functions of 14-3-3 may then be to provide a scaffold on which other proteins interact. Heterodimer formation would also provide a mechanism for specificity in the formation of a complex between two different protein kinases and 14-3-3. The structure of 14-3-3 presented here provides a framework on which these proposals may be tested in detail. □

Crystal structure of the zeta isoform of the 14-3-3 protein

Dong Liu, Jadwiga Bienkowska, Carlo Petosa, R. John Collier*, Haiyan Fu† & Robert Liddington

Dana-Farber Cancer Institute and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

* Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

† Department of Pharmacology, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA

THE 14-3-3 family of proteins have recently been identified as regulatory elements in intracellular signalling pathways¹: 14-3-3 proteins bind to oncogene and proto-oncogene products, including c-Raf-1 (refs 2–5), c-Bcr (ref. 6) and polyomavirus middle-T antigen⁷; overexpression of 14-3-3 activates Raf kinase in yeast^{2,3} and induces meiotic maturation in *Xenopus* oocytes⁵. Here we report the crystal structure of the major isoform of mammalian 14-3-3 proteins at 2.9 Å resolution. Each subunit of the dimeric protein consists of a bundle of nine antiparallel helices that form a palisade around an amphipathic groove. The groove is large enough to accommodate a tenth helix, and we propose that binding to an amphipathic helix represents a general mechanism for the interaction of 14-3-3 with diverse cellular proteins. The residues in the dimer interface and the putative ligand-binding surface are invariant among vertebrates, yeast and plants, suggesting a conservation of structure and function throughout the 14-3-3 family.

We crystallized a recombinant 14-3-3 protein (ζ isoform) expressed in *Escherichia coli*, and solved its structure by using two heavy-atom derivatives, anomalous scattering and non-crystallographic symmetry phase refinement (Table 1). The crystals contain two dimers in the asymmetric unit arranged with 222 point symmetry. A complete atomic model for the 245-residue protein has been built, with the exception of a flexible loop formed by residues 203–216 between helices $\alpha 8$ and $\alpha 9$ (Fig. 1).

Each monomer comprises a single domain of nine α -helices ($\alpha 1$ to $\alpha 9$) that are organized in a simple antiparallel fashion with mostly short loops between them (Fig. 2). Helices $\alpha 3$ and $\alpha 4$ are the longest (30 or 33 residues), spanning the length of the monomer (55 Å), and adopting a coiled-coil structure that extends 35 Å beyond $\alpha 1$ and $\alpha 2$. Helix $\alpha 3$ is bent in the middle at a conserved glycine (G53). The four amino-terminal α -helices form a plough-shaped module that packs against its symmetry-

Received 24 March; accepted 5 May 1995.

1. Aitken, A. *Trends biochem. Sci.* **20**, 95–97 (1995).
2. Robinson, K. et al. *Biochem. J.* **299**, 853–861 (1994).
3. Ichimura, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7084–7088 (1988).
4. Burbelo, P. D. & Hall, A. *Curr. Biol.* **5**, 95–96 (1995).
5. Freed, E., Symons, M., McDonald, S. G., McCormick, F. & Ruggieri, R. *Science* **265**, 1713–1715 (1994).
6. Fu, H. et al. *Science* **266**, 126–128 (1994).
7. Irie, K. et al. *Science* **265**, 1716–1719 (1994).
8. Franti, W. J. et al. *Nature* **371**, 612–614 (1994).
9. Reuther, G. W., Fu, H., Cripe, L. D., Collier, R. J. & Pendergast, A. M. *Science* **266**, 129–133 (1994).
10. Tokar, A., Ellis, C. A., Sellers, L. A. & Aitken, A. *Eur. J. Biochem.* **191**, 421–429 (1990).
11. Aitken, A., Ellis, C. A., Harris, A., Sellers, L. A. & Tokar, A. *Nature* **344**, 594 (1990).
12. Mochly-Rosen, D., Khaner, H., Lopez, J. & Smith, B. L. *J. biol. Chem.* **268**, 14866–14868 (1991).
13. Roth, D., Morgan, A. & Burgoyne, R. D. *FEBS Lett.* **320**, 207–210 (1993).
14. Aitken, A., Howell, S., Jones, D., Madrazo, J. & Soneji, Y. *J. biol. Chem.* **270**, 5706–5709 (1995).
15. Jones, D. H. A. et al. *J. molec. Biol.* **245**, 375–384 (1995).
16. Rodgers, D. R. *Structure* **2**, 1135–1140 (1994).
17. Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC, Warrington, 1991).
18. Collaborative Computing Project no. 4. *Acta crystallogr.* **D50**, 760–763 (1994).
19. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 39–48 (SERC, Warrington, 1991).
20. Madden, D. R. thesis, Harvard Univ. (1993).
21. Jones, T. A. *Meth. Enzym.* **115**, 157–171 (1985).
22. Brünger, A. T. *XPLOR* version 3.0. (Yale Univ., New Haven, 1992).
23. Carson, M. J. *appl. Crystallogr.* **24**, 958–961 (1991).
24. Nicholls, A., Sharp, K. A. & Honig, B. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

ACKNOWLEDGEMENTS. We thank B. Rassmussen and C. I. Brändén at ESRF Grenoble, and P. Rizkallah and P. Lindley at SRS Daresbury for access to synchrotron facilities at short notice; K. Henrick for help with software; G. Davies, D. Trentham and J. Skehel for critical reading of this manuscript; M. He for assistance with crystallization; and P. Nielsen for the original τ 14-3-3 complementary DNA.

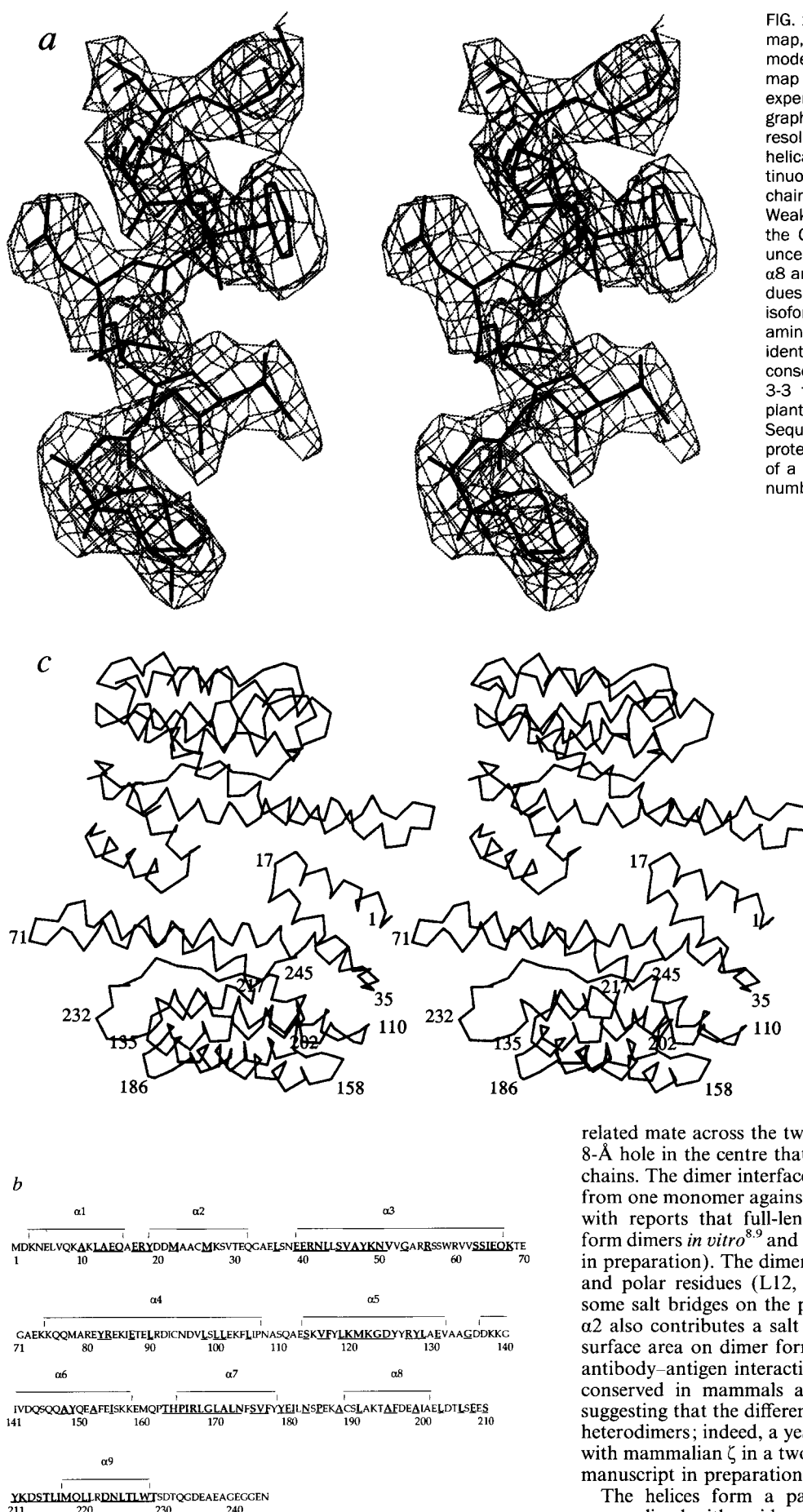


FIG. 1 **a**, Stereo image of the electron density map, superimposed on part of the final refined model of the $\alpha 3$ helix (residues 45 to 55). The map was generated from model-independent experimental phases after 4-fold non-crystallographic symmetry phase refinement at 3.2-Å resolution, and is contoured at 1σ . All of the helical segments are well ordered, and continuous electron density exists for the main-chain atoms and most of the side chains. Weaker density exists for the loops, particularly the C-terminal loop, for which the register is uncertain (± 2 residues), and the loop between $\alpha 8$ and $\alpha 9$, for which no density exists for residues 203 to 216. **b**, Sequence of the 14-3-3 ζ isoform with α -helical segments indicated. The amino-acid sequence of bovine 14-3-3 ζ is identical to human. Residues that are strictly conserved in the known members of the 14-3-3 family, including vertebrates, yeast and plants, are shown in bold underlined type. Sequences were taken from the SWISS-PROT protein sequence data bank. **c**, Stereo Ca plot of a dimer of 14-3-3, with occasional residue numbering.

related mate across the two-fold axis of the dimer, leaving a 6–8-Å hole in the centre that is lined with charged and polar side chains. The dimer interface is formed by the packing of helix $\alpha 1$ from one monomer against $\alpha 3$ and $\alpha 4$ from the other, consistent with reports that full-length and C-terminal-deleted proteins form dimers *in vitro*^{8,9} and *in vivo* (S. Levy and H. F., manuscript in preparation). The dimer interface buries several hydrophobic and polar residues (L12, A16, S58, V62, I65 and Y82), with some salt bridges on the periphery (R18–E89; E5–K74). Helix $\alpha 2$ also contributes a salt bridge (D21–K85). The total buried surface area on dimer formation is 620 Å², similar to a typical antibody–antigen interaction. The interface residues are strictly conserved in mammals and yeast (the salt bridges less so), suggesting that the different 14-3-3 isoforms would readily form heterodimers; indeed, a yeast 14-3-3 protein forms heterodimers with mammalian ζ in a two-hybrid system (L. Zhang and H. F., manuscript in preparation).

The helices form a palisade that defines an amphipathic groove lined with residues from helices $\alpha 3$, $\alpha 5$, $\alpha 7$ and $\alpha 9$ (Fig.

TABLE 1 Summary of crystallographic analysis

Crystals	Space group: $P6_5$		$a = 94.9 \text{ \AA}$; $c = 236.7 \text{ \AA}$		4 molecules/asymmetric unit			56% solvent content		
Data set	λ (\AA)	$d_{\min}$ (\AA)	Unique reflections	Completeness (%)	$\langle I \rangle / \sigma(I)$	R_{sym} (%)	R_{iso} (%)	No. of sites	Phasing power	FOM
Native	1.010	2.8	28,237	96.4	11.6	7.4				0.60
Hg λ 1	0.950	2.9	50,449	99.7	9.2	6.9	17.9	12	2.6	
Hg λ 2	1.010	2.8	54,021	95.4	7.5	7.4	15.7	12	2.1	
Se-Met	0.979	3.6	22,695	88.9	11.6	6.8	17.1	24	0.4	
Refinement statistics										
Resolution (\AA)		C_{initial}	C_{final}	No. of atoms	R_{cryst} (%)	R.m.s. bond length (\AA)		R.m.s. bond angle (deg.)		
8.0–2.9		0.39	0.91	1,847	21.4	0.019		3.9		

$R_{\text{sym}} = \Sigma |I - \langle I \rangle| / \Sigma I$, where I is observed intensity, $\langle I \rangle$ is average intensity from multiple observations of symmetry-related reflections. $R_{\text{iso}} = \Sigma ||F_{\text{der}}| - |F_{\text{nat}}|| / \Sigma |F_{\text{nat}}|$, where $|F_{\text{nat}}|$ is the protein structure factor amplitude and $|F_{\text{der}}|$ the heavy-atom derivative structure factor amplitude. Phasing power is defined as the root-mean-square ($|F_{\text{H}}|/E$), where $|F_{\text{H}}|$ is the heavy-atom structure factor amplitude and E the residual lack of closure. FOM is the mean figure of merit. C_{initial} and C_{final} are correlation coefficients of the electron density for the non-crystallographic symmetry operators. Expression and purification of recombinant 14-3-3 from bovine brain was described previously¹⁵. The Se-Met form was produced as in ref 16. Crystals were grown at 4 °C by vapour diffusion in hanging drops (5–10 mg ml⁻¹ protein, reservoir: 20–24% PEG 3500, 100 mM Tris-HCl, pH 8.5, 10 mM MgCl₂, 1 mM NiCl₂). The Hg derivative was produced by crystallization in the presence of 4 mM thimerosal. Crystals were transferred in 3 steps into (reservoir + 4% PEG 3500 + 15% ethylene glycol) before flash-freezing at 100 K. All data were collected at beamline X12C (NSLS, Brookhaven, New York). The native and mercurated crystal were oriented such that Bijvoet pairs were collected on the same or adjacent images. Derivative data sets were collected at a remote point (Hg λ 1) and at the inflection point (Hg λ 2) of the Hg absorption edge and at the Se absorption edge for the Se-Met crystals. Data were processed by DENZO and SCALEPACK (Z. Otwinowski). One Hg position was determined by inspection of an isomorphous difference Patterson; 11 further sites were found from difference Fouriers. 24 Se positions were readily obtained from an anomalous difference Fourier using the mercury phases. Heavy-atom refinement utilized HEAVY¹⁷ and MLPHARE (Z. Otwinowski), using the two isomorphous and three anomalous data sets, and tabulated values of f' and f'' . Non-crystallographic symmetry operators were determined initially from the Hg atom positions. An initial mask was created from an averaged map and O (ref. 18), and phases refined using the 4-fold non-crystallographic symmetry and solvent flattening (program RAVE; G. J. Kleywegt and T. A. Jones). A preliminary model was built into this map by using FRODO¹⁹, which provided an improved mask for a further round of phase improvement. A model containing most of the α -helices was subjected to least-squares and simulated-annealing refinement²⁰ followed by phase recombination. Loops were built into this improved map, followed by further rounds of refinement.

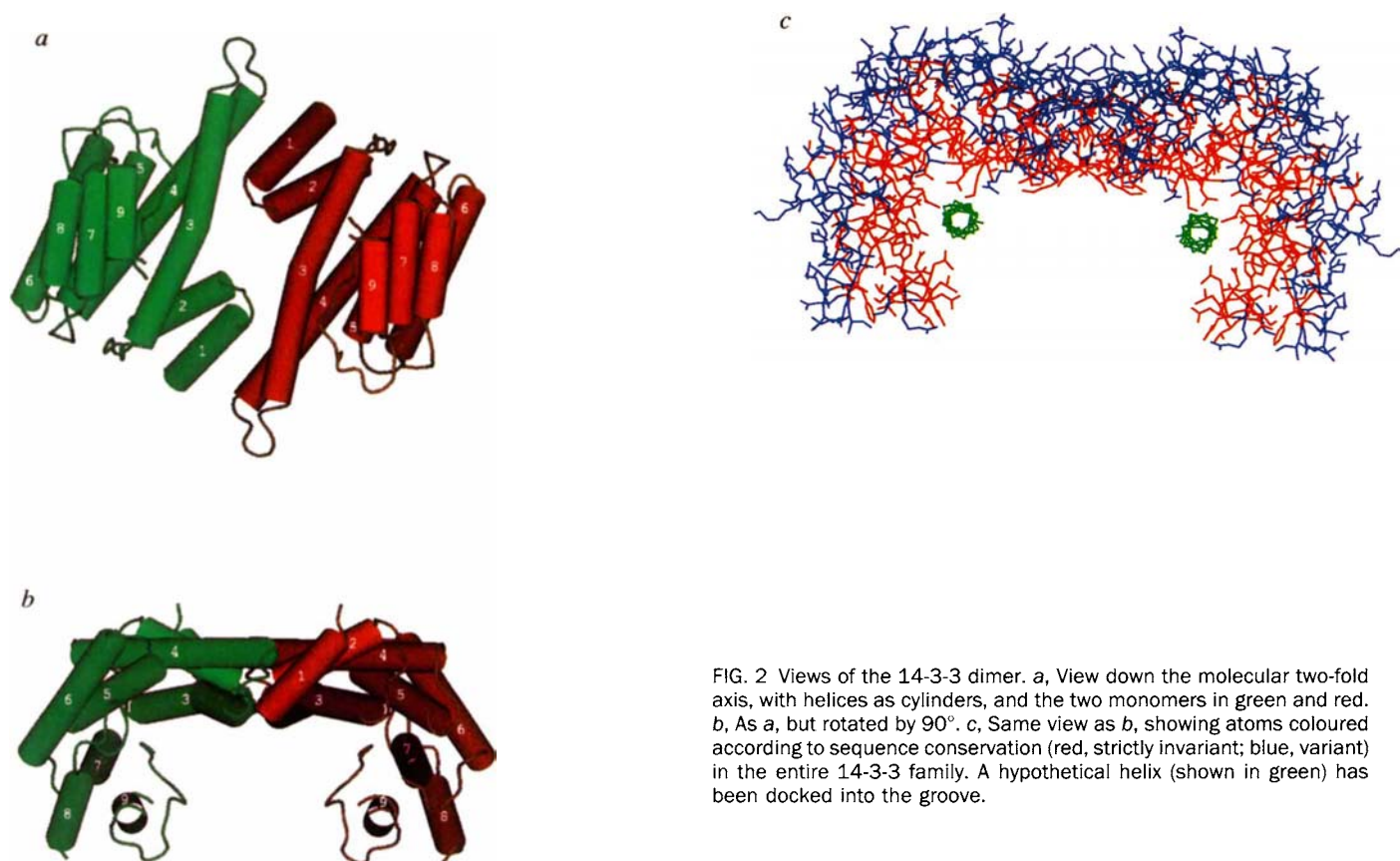


FIG. 2 Views of the 14-3-3 dimer. *a*, View down the molecular two-fold axis, with helices as cylinders, and the two monomers in green and red. *b*, As *a*, but rotated by 90°. *c*, Same view as *b*, showing atoms coloured according to sequence conservation (red, strictly invariant; blue, variant) in the entire 14-3-3 family. A hypothetical helix (shown in green) has been docked into the groove.

3). On one side of the groove, $\alpha 7$ and $\alpha 9$ present a hydrophobic surface that includes four exposed leucine side chains (L172, L216, L220 and L227). On the other side, $\alpha 3$ presents three basic side chains (K49, R56 and R60) and $\alpha 5$ presents polar and charged groups (K120, D124, Y128 and R127). The three arginine residues form a basic cluster near the top of the groove. In our crystal structure, the final 15 residues of the protein (231 to 245) form a poorly ordered loop, the end of which rests on the surface of the groove, partly obscuring the basic cluster.

The sequences of 14-3-3 proteins are highly conserved from vertebrates to plants. We examined the 17 sequences of currently known 14-3-3 family members, and found that 104 residues were strictly invariant. When these are displayed on our three-dimensional model, the molecule partitions sharply into a strictly invariant surface (the concave surface that includes the lining of the groove), and a variable region (the outer convex surface), suggesting that the groove may have a conserved function (Fig.

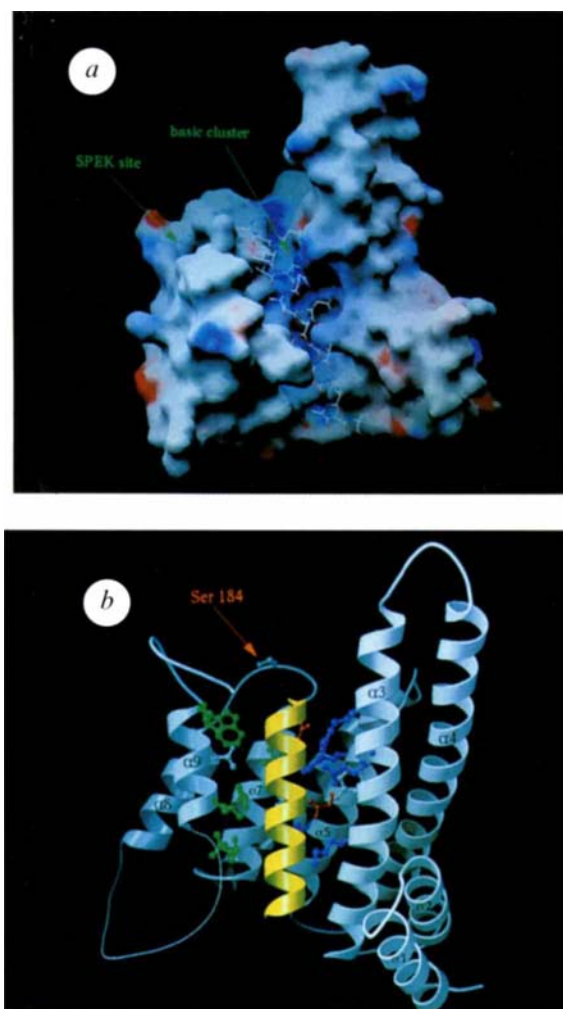


FIG. 3 a, Surface-charge representation of a monomer of 14-3-3 calculated with the program GRASP²¹, after removal of the C-terminal loop (231 to 245). The view is similar to that of the green monomer in Fig. 2a. A hypothetical amphipathic helix has been docked into the groove. The basic cluster and phosphorylation site are indicated. b, Schematic representation of a monomer of 14-3-3, with side chains lining the amphipathic groove indicated. Side chains are coloured according to type: hydrophobic (green), basic (blue), acidic (red) and polar (grey). The hydrophobic side of the groove is formed by the exposed side chains of L172, V176, L216, I217, L220, L227 and W228. The right side contains charged and polar residues, including a cluster of arginines (R56, R60 and R127). A schematic model of an α -helix (main chain only) docked into the groove is shown in yellow. The location of the phosphorylation site, S184, is indicated. The disordered loop formed by residues 203 to 216 is shown schematically as a thin ribbon.

2c). The carboxy-terminal loop sequence is not invariant in the 14-3-3 family, but it is always acidic, and in our crystal structure it interacts with the cluster of basic residues at the top of the groove. The role of this loop may be regulatory, or it may help to stabilize the unliganded structure.

We propose that the C-terminal loop is ejected when 14-3-3 binds ligand, exposing the surface of the groove, whereupon it binds an amphipathic helix from the target protein. We docked a model helix into the groove after removal of the C-terminal loop (Fig. 3); an amphipathic helix containing leucine and asparagine residues fits comfortably, making a leucine-zipper type of contact with $\alpha 9$ and $\alpha 7$, and polar contacts with the side chains of $\alpha 3$ and $\alpha 5$. Little information is available on the sequence motifs of proteins that bind to 14-3-3, but it has recently been shown that 14-3-3 ζ possesses chaperone-like properties in the mitochondrial import process, and that this process is mediated by binding to a variety of amphipathic signal sequences¹⁰, in support of our hypothesis.

How is ligand binding regulated? Several reports have demonstrated or implicated phosphorylation of the target protein in high-affinity binding^{2,4,6,11}. Arginine and lysine residues are usually involved in phosphoaminoacid binding sites¹² and there is an appropriate cluster on the basic face of the groove of 14-3-3 (R56, R60, R127 and K120) that might form such a binding site; we are currently testing this hypothesis. Phosphorylation of 14-3-3 may also regulate activity: the ζ isoform (but not τ , η or γ) is phosphorylated at a proline-directed kinase consensus site (SPEK), which increases its affinity for protein kinase C (ref. 13); this site is located at the beginning of helix $\alpha 8$, near the top of the groove (Fig. 3), where it could affect ligand binding directly.

Our results suggest that 14-3-3 proteins have a ligand-binding surface that has been absolutely conserved throughout evolution; indeed, a yeast 14-3-3 protein is able to activate mammalian Raf-mediated signalling³. Our model begs comparison with the mechanism of action proposed for calmodulin, another highly conserved protein that can bind a range of amphipathic helices and active proteins involved in signal transduction pathways¹⁴. The functional significance of 14-3-3 dimerization is unknown, but perhaps the simultaneous binding of helices from two target proteins can induce both homo- and heterodimer formation as a step in the signal transduction process. The precise role of 14-3-3 in cellular regulation awaits further biochemical and genetic analysis; our crystal structure provides a molecular basis for future exploration. □

Received 19 April; accepted 6 June 1995.

- Burbelo, P. & Hall, A. *Curr. Biol.* **5**, 95–96 (1995).
- Freed, E., Symons, M., Macdonald, S. G., McCormick, F. & Ruggieri, R. *Science* **265**, 1713–1716 (1994).
- Irie, K. et al. *Science* **265**, 1716–1719 (1994).
- Fu, H. et al. *Science* **266**, 126–129 (1994).
- Fantl, W. et al. *Nature* **371**, 612–614 (1994).
- Reuther, G. W., Fu, H., Cripe, L. D., Collier, R. J. & Pendergast, A. M. *Science* **266**, 129–133 (1994).
- Pallas, D. C. et al. *Science* **265**, 535–537 (1994).
- Boston, P. F., Jackson, P., Kynoch, P. A. M. & Thompson, R. J. *J. Neurochem.* **38**, 1466–1474 (1982).
- Jones, D. et al. *J. molec. Biol.* **245**, 375–384 (1995).
- Adam, R. et al. *J. Biochem., Tokyo* **116**, 416–425 (1994).
- Furukawa, Y. et al. *Biochem. biophys. Res. Commun.* **194**, 144–149 (1993).
- Waksman, G. et al. *Nature* **358**, 646–653 (1992).
- Aitken, A., Howell, S., Jones, D., Madrazo, J. & Patel, Y. *J. biol. Chem.* **270**, 5706–5709 (1995).
- Finn, B. & Forsén, S. *Structure* **3**, 7–11 (1995).
- Fu, H., Coburn, J. & Collier, R. J. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2320–2324 (1993).
- Harrison, C. J., Bohm, A. A. & Nelson, H. C. M. *Science* **263**, 224–227 (1994).
- Terwilliger, T. & Eisenberg, D. *Acta crystallogr.* **A39**, 813–817 (1983).
- Jones, T. et al. *Acta crystallogr.* **A47**, 110–119 (1991).
- Jones, T. A. *Meth. Enzym.* **115**, 157–171 (1985).
- Brünger, A. *XPLOR Version 3.1: A system for X-Ray Crystallography and NMR* (Yale University Press, New Haven, CN, 1992).
- Nicholls, A., Sharp, K. A. & Honig, B. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

ACKNOWLEDGEMENTS. We thank R. Sweet and his staff at beamline X12C of the Brookhaven National Light Source for assistance in data collection; the facility is supported by the U.S. Department of Energy, Office of Health and Environmental Research, and by the NSF. R. L. was supported by the NIH and Sandoz Pharmaceuticals; H. F. was supported in part by a grant from the American Cancer Society.

A column on chromatography

This week's review checks in with a gradient delivery system, an HPLC detector, protein separation columns, a tentacle-gel medium for affinity chromatography and a variety of data handling systems.

HPLC TECHNOLOGY now offers Snap-Cap cartridges, which are **solid-phase extraction devices** that are useful for the clean up of samples before analysis by chromatography (*Reader Service No. 100*). This new cartridge design incorporates two important features: a faster flow rate and less dead volume. It is moulded of low-extractable, solvent-resistant polypropylene resin. The inlet end of the cartridge has a female luer taper for a positive leak-free connection to a syringe. The outlet is a standard male luer taper, which permits connection to a syringe. There are 15 high-quality, chemically bonded sorbent beds available.

The GP40 gradient pump, from Dionex, is a **dual-processor-controlled, solvent-gradient delivery system** designed to blend and pump mixtures of up to four mobile phase components at precisely controlled flow rates (*Reader Service No. 101*). The pump's dual processors — a microprocessor for module common functions and a digital signal processor dedicated to full-time pump control — ensure tight regulation for superior flow accuracy and excellent flow stability for low noise. The selected mobile phase composition may be delivered either isocratically or as a linear or exponential gradient. The computer-aided design of the flow path is said to provide extremely small delay volumes for accurate, sharp gradients and advanced mixer designs are said to ensure smooth composition. The GP40 can be configured for either standard bore or microbore flow rates. Passivated stainless steel and polyetherether ketone (PEEK) flow configurations are available.

Sepragen has released its next generation products and technology for the process separations market with the introduction of the QuantaSep 1000 **computer-controlled, process-chromatography workstation** (*Reader Service No. 102*). This is an integrated chromatography system enabling complete automation of liquid chromatography operations, including equilibration, sample application, washing, reproducible gradient formulation, and selective sample fraction collection on three columns to occur automatically by pre-programming all operations. Continuous monitoring and control of each step in the purification process is



Sepragen's QuantaSep 1000.

performed with on-line UV, pH, and conductivity sensors, and pressure, air, and leak detectors. Monitored values are used to make decisions automatically to the process. The software allows for the continuous recording, storage and display of all data generated throughout the entire process. The QuantaSep 1000 is designed to meet all GMP requirements for documentation, validation, sanitation, data storage, and password security.

ESA has introduced a new **stand-alone HPLC detector** that utilizes an array of coulometric, electrochemical sensors for detecting ultra-low (nanogram to femtogram) levels of analytes in complex samples (*Reader Service No. 103*). This should prove useful for the areas of cancer, environmental, biological, food and pharmaceutical research. By resolving electrochemically active compounds across an array of up to 16 sensors, the model 5600 CoulArray Detector can separate substances that co-elute and would thus be missed by other detection methods. The result is a unique "fingerprint" of the material being analysed. The manufacturer states that the model 5600 is the first detector that makes gradient HPLC practical with electrochemical detection. The data system automatically corrects for the electrochemical baseline changes caused by the gradient. Completely

integrated gradient and isocratic system configurations are available.

■ Data systems

Shimadzu Scientific Instruments announces its latest **PC-based chromatography data acquisition systems software** (*Reader Service No. 104*). The CLASS-VP features new data-processing and compliance tools, automated batch feedback and control of more instruments. Shimadzu's chromatography data system is a unification of both Shimadzu's EZChrom and CLASS software. It is a multi-tasking chromatography data acquisition and instrument-control system operating in a Microsoft Windows environment. The data manipulation package has been refined with additional calibration types, integration events including the Shimadzu Chromatopac integration algorithm with automatic peak integration. New GLP features include storing results with the data and method, file overwrite protection and intelligent batch table with decision-making capability. Custom report features include WYSIWYG support allowing up to four chromatograms and reports per page, custom formatting of the report including user-defined custom parameters. The system takes control of multiple Shimadzu LCs or GCs or a combination of both to the next level in automation.

The latest upgrade of the ATI Unicam 4880 **chromatography data system** incorporates a 'continuous run' option for improved automation and increased productivity (*Reader Service No. 105*). This provides a new option to the Windows-based data system and features a number of automatic turnkey analyser systems. Comprehensive graphic facilities include multi-channel,



ATI Unicam's 4880 data system.

on-screen displays with zoom boxes and full-colour chromatogram overlay. These are designed to facilitate fast and effective data review. Data are easily exported to powerful spreadsheets for application-specific calculations and customized reporting.

Jones Chromatography has launched the new JCL6000, a **chromatography data system** for Windows (*Reader Service No. 106*). This is designed to be a low-cost, icon-driven system, which can handle up to 100 samples in a single method, as well as up to 50 standards. An unlimited number of chromatograms can be compared in the overlay feature. The system has a full range of peak-integration options, including timed integration events controlling multiple sets of integration parameters, forced baselines and manual baseline adjustments. The Windows operating environment allows 'cut-and-paste' of chromatograms to other Windows packages, enabling the user to produce professional quality reports.

■ Columns

Fluorescence labelling is combined with new **HPLC columns** in two new kits from Oxford Glycosystems (*Reader Service No. 107*). Used together, the Signal labelling kit and the GlycoSep HPLC column set provide high-resolution HPLC analysis of oligosaccharides at the picomole level. Signal uses 2-aminobenzamide (2-AB) to label oligosaccharides at their reducing terminus; from as little as 25 pmol up to 50 nmol of sample oligosaccharide — 2-AB labelled

from glycoproteins in the low picomole range. A standard HPLC system with a fluorescence detector (excitation 300 nm and emission 420 nm) is required.

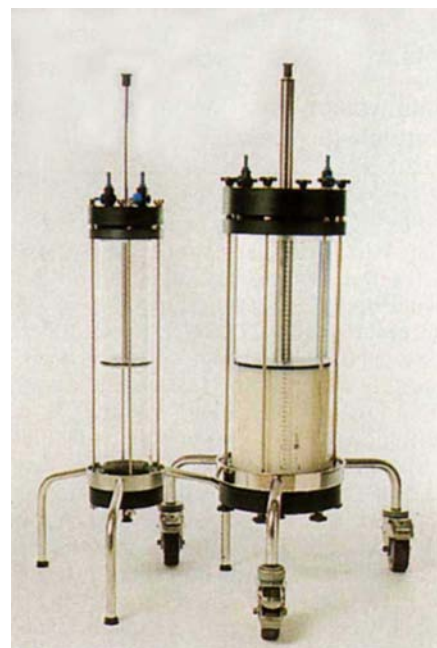
Chrompack has introduced a series of **low-bleed GC-MS columns**, which are each tested individually for standard performance and low bleeding performance (*Reader Service No. 108*). The capacity and selectivity criteria are said to be virtually identical to their all-around CP-Sil 5 CB, but with significantly lower bleeding. The resulting CP-Sil 5 CB-MS column is available in two lengths (25 m and 50 m), two internal diameters and four film thicknesses, all available from stock.

The new Hypersil Duet columns from Shandon HPLC have been designed for the **simultaneous separation of hydrophobic and ionic compounds**, thereby overcoming the problem of pre-separation of different types of analyte in a sample (*Reader Service No. 109*). Whereas HPLC columns are usually intended to separate only one type of analyte in a given chromatographic run, these columns being composed of a mixture of ion-exchange and hydrophobic media, are said to permit the separation of both hydrophobic and ionic species in a single analysis. As the chromatographic properties are intermediate between each phase, the retention of ionic species can be accomplished by a change in pH with no consequence to the retention of hydrophobic molecules as long as the solubility is not affected. These columns are available as mixtures of C18 with a strong cation exchanger and C18 with a strong anion exchanger in 250 mm × 4.6 mm geometries.

A new brochure detailing the extensive range of **HPLC-certified sample preparation filters** by Micron Separations is now available from Porvair Filtronics (*Reader Service No. 110*). In addition to complete ordering information, the brochure includes photographs showing examples from each of the six size- and volume-based product

categories, each of which offer a choice in membrane and pore size. It also includes an application guide to aid selection of the correct filter membrane.

Pharmacia Biotech has introduced a new **INdEX glass column**, which is designed to meet the requirements of both diagnostic production and the



INdEX columns from Pharmacia Biotech.

development laboratory (*Reader Service No. 111*). This new general-purpose column is said to be easy to work with and simple to operate. Column packing is said to be a 5-min operation with the hydraulically adjustable adaptor, which uses a novel and reproducible packing method. Furthermore, the adaptor is adjustable, which makes a wide range of bed heights and volumes possible. Column tubes are available with inner diameters of 100 and 200 mm and lengths of 500 and 950 mm. They are designed for use with high flow rates and are pressure rated up to 300 kPa.

Chiral Technologies now offer the Chiralcel OD **chiral HPLC columns** (*Reader Service No. 112*). These columns have a cellulose carbamate/silica gel adsorbent, which offers broad functionality for separating chiral molecules and will separate racemic mixtures in quantities ranging from milligrams to grams. The column is compatible with standard HPLC instrumentation and is available in a range of sizes. In addition to 2-cm inner-diameter columns, there are 0.46 cm (analytical), 1 cm, 5 cm, 10 cm and 20 cm inner-diameter columns available. Columns 5-cm diameter and larger are equipped with integral water jackets and are mounted on wheeled carts for easy mobility.

Waters introduces a new product for **high-purity protein separations**, the Protein-Pak HR glass column Easy-Pak kits (*Reader Service No. 113*). These kits are designed to provide protein chemists with high-purity protein separations at a reduced cost. The kit comes complete



Oxford Glycosystems' GlycoSep HPLC column.

glycans can be prepared in high yield (>85 per cent) in about 2 h. The labelled oligosaccharides can then be analysed by two-dimensional HPLC using a charge-based separation on GlycoSep C and a hydrophobicity-based separation on GlycoSep H. In tandem, these columns are said to separate oligosaccharide mixtures released



Waters' Protein-Pak HR Easy-Pak.

with a 5 mm × 100 mm 'advanced purification' column, 5 grams of packing material in 15-μm high-resolution SP, Q, DEAE, or CM chemistries, as well as self-pack accessories. The refill kit contains 5 g of packing material of choice.

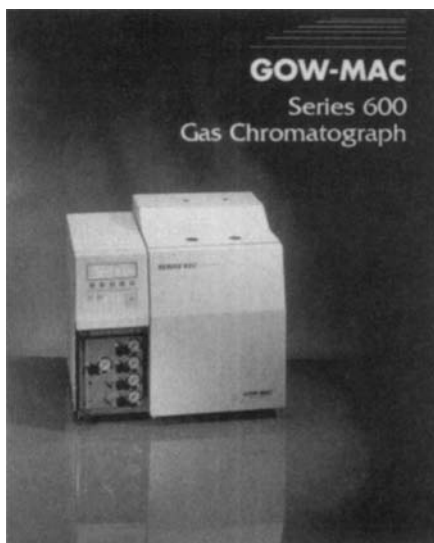
■ LC/GC

Now Varian's fast GC technology has been engineered into its 3000 series Star gas chromatographs and combined with several other time-saving features to bring users a fast, versatile system (Reader Service No. 114). Called the Star+ series, this instrument is designed to minimize sample handling by streamlining the analyte isolation process. By incorporating a solid-phase, micro-extraction process into the system, the manufacturer says that it has reduced the total analysis time by 50 per cent for several time-consuming applications. This solvent-free process utilizes microfibres to extract the desired analyte and inject it directly into the GC. The company has made the system fully compatible with FID, ECD and PID detectors so that these time-saving benefits are available for a broader range of applications.

Fisons Instruments has developed a fully automatic, on-line-coupled LC/GC with the Dualchrom 3000 (Reader Service No. 115). The system has been designed for efficient pre-separation, thorough clean up and

effective sample enrichment. This is achieved by using liquid chromatography, which eliminates most of the disadvantages associated with off-line methods such as manipulation and solute losses. The automated sample preparation stage may be integrated into the final chromatographic analysis for increased productivity and reduced cost.

Gow-Mac Instruments is offering a brochure on the new series 600 GC, a programmable, research-grade gas chromatograph (Reader Service No. 116). The instrument is capable of uti-



Series 600 gas chromatograph.

lizing eight interchangeable detectors: TCD, FID, ECD, PID, NPD, ADD, FPD, and DID. Two detectors may be installed simultaneously with packed or capillary columns and split/splitless, direct on-column, gas-sample injection may be accommodated.

■ Media

Merck has developed Fractogel EMD activated, polymer-based, tentacle gels for affinity chromatography (Reader Service No. 117). The manufacturer states that a wide range of applications with different coupling conditions for immobilizing ligands can be performed using the advantages of the tentacle technique. The gel itself is manufactured to be chemically stable and allow high flow rates to be applied. The hydrophobic character of the tentacle spacers reduces nonspecific hydrophobic interactions between proteins and the matrix. The

azlactone medium is recommended for the immobilization of proteins under physiological conditions; and the epoxy medium represents a matrix for efficient binding of a variety of compounds.

Phenomenex has created Prodigy to be an ultra-pure, high efficiency HPLC packing material available in both C8 and ODS(2) bonded phases (Reader Service No. 118). It is designed to eliminate the need for mobile phase modifiers such as ion-pairing reagents and competing amines while producing high-efficiency symmetrical peaks, even with the most polar acidic or basic compounds. Minimal silanol activity provides improved chromatography of organic amines and acids, amphoteric as well as neutral compounds. Prodigy is said to be a cost-effective alternative to various other packing materials. All packed columns are available in sizes ranging from 1 mm to 100 mm inner diameter. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

DNA SEQUENCING

Using an ABI 373
Automated Sequencing
System

£45 per run

- PCR products
- M13 inserts
- plasmid inserts
- single stranded DNA
- double stranded DNA
- 300-400 bases of sequence

Please contact us for more information

Molecular Medicine Unit

King's College School of Medicine & Dentistry
Tel: 071-346 3126 Fax: 071-733 3877

Reader Service No.11

XX IDT
INTEGRATED DNA TECHNOLOGIES, INC.

Custom DNA Synthesis

\$1.95 / Base

200 nmole Scale

No Set-up Fee

Column Purified

24 Hour Shipping

Q.C. by PAGE on Every Oligo

E-mail: orders@bbs.idtdna.com

Fax: 1-319-645-2921

Phone: 1-800-328-2661

Major Credit Cards Accepted

Reader Service No.3



Fisons' Dualchrom 3000 on-line coupled LC/GC.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of nonspecialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £700 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at Porters South, 4 Crinan Street, London N1 9SQ, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.